

Revista

ESTUDIOS DE LINGÜÍSTICA DEL ESPAÑOL

Estudios de Lingüística del Español 45 (2022)

Volumen monográfico:

Metodologías lingüísticas: de los datos empíricos a la teoría del lenguaje

Coordinadoras: Andrea Ariño-Bizarro y Natalia López-Cortés

© *Estudios de Lingüística del Español* 2022. Reservados todos los derechos.
ISSN: 1139-8736
<https://www.infoling.org/elies/>

Estudios de Lingüística del Español (ISSN: 1139-8736) es una revista especializada en temas relacionados con la lingüística hispánica.

Estudios de Lingüística del Español (ELiEs) publica un volumen al año.

El envío de contribuciones para la revista ELiEs se debe realizar por correo-e dentro de los plazos establecidos en las correspondientes peticiones de contribuciones, que se anuncian en la lista de distribución *Infoling* (<https://www.infoling.org/>).

La información necesaria para el envío de originales se encuentra en la página de *Normas de edición* de la web de ELiEs, en <https://infoling.org/elies/?p=normas-edicion>.

En su versión en línea, que se encuentra en <https://www.infoling.org/elies>, *Estudios de Lingüística del Español* se publica conjuntamente con *Infoling Revista*, un boletín informativo permanentemente actualizado, que recoge las reseñas y la información sobre publicaciones, congresos, ofertas de trabajo, etc., que se difunden en *Infoling* por correo-e y en las redes sociales.

ELiEs se edita con la ayuda de la Universitat Autònoma de Barcelona, la Universität Bern y la Ruhr-Universität Bochum.

Estudios de Lingüística del Español está indizada en las siguientes bases de datos: [Clasificación Integrada de Revistas Científicas \(CIRC, 2ª edición 2011/12\)](#), [Dialnet](#), [Dulcinea](#), [Google Scholar](#), [Latindex](#), [Linguistic Bibliography](#), [Matriu d'Informació per a l'Anàlisi de Revistes \(MIAR\)](#), [RACO](#), [ROAD \(Directory of Open Access Scholarly Resources\)](#), [SCOPUS](#), [WorldCat](#), [Zeitschriftendatenbank \(ZBD\)](#).

© *Estudios de Lingüística del Español* 2022. Reservados todos los derechos.

ISSN: 1139-8736

<https://www.infoling.org/elies/>

Equipo editorial

Dirección

Carlos Subirats Rüggeberg (Universidad Autónoma de Barcelona)

María Luisa Calero Vaquera (Universidad de Córdoba, España)

Yvette Bürki (Universität Bern)

Laura Morgenthaler García (Ruhr-Universität Bochum)

Desarrollo y programación

Marc Ortega Gil (Universidad Autónoma de Barcelona)

Compaginación Matthias Lehner

Comité científico

Alexandra Álvarez (Universidad de los Andes, Venezuela)

Valerio Báez San José (Universidad Carlos III)

Paola Bentivoglio (Universidad Central de Venezuela)

Yvette Bürki (Universität Bern)

Teresa Cabré (Universitat Pompeu Fabra)

Miguel Casas Gómez (Universidad de Cádiz)

Luis Cortés (Universidad de Almería)

Mar Cruz Piñol (Universidad de Barcelona)

Adolfo Elizaincín (Universidad de la República, Montevideo)

Miguel Ángel Esparza (Universidad Rey Juan Carlos)

Milagros Fernández Pérez (Universidad de Santiago de Compostela)

Marie-Claude L'Homme (Université de Montréal)

Covadonga López Alonso (Universidad Complutense de Madrid)

Juan de Dios Luque (Universidad de Granada)

Francisco A. Marcos Marín (University of Texas San Antonio)

Salvio Martín Menéndez (Universidad de Buenos Aires / CONICET)

Emma Martinell (Universidad de Barcelona)

Hans-Josef Niederehe (Universität Trier)

Ramón Sarmiento (Universidad Rey Juan Carlos)

Maite Taboada (Simon Fraser University, Canadá)

José del Valle (City University of New York)

Alba Valencia (Universidad de Chile)

Isabel Verdaguer (Universidad de Barcelona)

Gerd Wotjak (Universität Leipzig)

Comité de edición

Julia Bernd (International Computer Science Institute)

Paloma Garrido Íñigo (Universidad Rey Juan Carlos)

Matthias Raab (Universidad de Barcelona)

Laura Romero (Universidad de Barcelona)

Reseñas

Coordinadora de reseñas: Marta Estévez Grossi (Leibniz Universität Hannover)

Coordinadora técnica de reseñas: Cristina Vela Delfa (Universidad de Valladolid)

Soporte informático: Marc Ortega (Universidad Autónoma de Barcelona)

Compaginación: Matthias Lehner (Universität Bern)

Estudios de Lingüística del Español 45 (2022)
VOLUMEN MONOGRÁFICO
Metodologías lingüísticas:
de los datos empíricos a la teoría del lenguaje
Coordinadoras: Andrea Ariño-Bizarro y Natalia López-Cortés

ÍNDICE

Introducción

ARIÑO-BIZARRO, ANDREA; LÓPEZ-CORTÉS, NATALIA. *La metodología lingüística en el siglo XXI*.....5

Artículos

BLOQUE I: METODOLOGÍAS LINGÜÍSTICAS DE LA TRADICIÓN

1. GARCÍA-MIGUEL, JOSÉ M. *Lingüística de corpus: de los datos textuales a la teoría lingüística*11
2. AURREKOETXEA, GOTZON. *Métodos de la dialectología cuantitativa*43
3. MORET OLIVER, MARÍA TERESA; GIRALT LATORRE, JAVIER. «*Sin documentos no hay historia*». *Aprovechamiento del corpus de documentos aragoneses en lengua catalana*.....71
4. IBARRETXE-ANTUÑANO, IRAIDE. *Trabajo de campo*103

BLOQUE II: METODOLOGÍAS LINGÜÍSTICAS DE LA CONTEMPORANEIDAD

5. IGOA, JOSÉ MANUEL. *Las tareas conductuales en la investigación sobre el procesamiento del lenguaje*.....133
6. MARRERO-AGUIAR, VICTORIA. *Aportación de las investigaciones con neuroimagen funcional a los estudios lingüísticos: algunas preguntas y respuestas*159
7. HARO, JUAN. *Una introducción al uso de los potenciales evocados en el estudio del lenguaje*185
8. ÁLVAREZ GARCÍA, ESTHER. *Lo que esconden tus ojos: la metodología eye-tracking aplicada al estudio del lenguaje*.....205
9. VALENZUELA, JAVIER. *El big data en los estudios del lenguaje*241
10. JUAN, NADÈGE; BÜRKI, YVETTE. *Entrevistas mediadas en sociolingüística cualitativa. La aplicación en la etnografía en línea*261

Reseñas

- CRESPO DEL RÍO, CLAUDIA. Reseña de Fuller, Janet M.; Leeman, Jennifer. 2020. *Speaking Spanish in the US. The Sociopolitics of Language*. Bristol: Multilingual Matters.....285
- GALLARDO RICHARDS, EMMA. Reseña de Dubert-García, Francisco; Míguez, Vítor; Sousa, Xulio. 2020. *Variedades lingüísticas en contacto na Península Ibérica*. Santiago de Compostela (España): Consello da Cultura Galega.....294
- DEL PRATO, JULIETA. Reseña de Carranza, Isolda E. 2020. *Narrativas interaccionales. Una mirada sociolingüística a la actividad de narrar en encuentros sociales*. Córdoba (Argentina): Editorial de la Facul.....299
- MALAVER, IRANIA. Reseña de Sáez, Daniel M.; Octavio de Toledo, Álvaro S. 2020. *Textos españoles de la primera mitad del siglo XVIII para la historia gramatical y discursiva. Vientos de arrastre y de cambio en la historia del español*. Madrid: Síntesis303
- GARCÍA AGÜERO, ALBA NALLELI. Reseña de Mancera, Ana; Pano, Ana. 2020. *La opinión pública en la red. Análisis pragmático de la voz de los ciudadanos*. Madrid / Frankfurt: Iberoamericana / Vervuert.308

RODRÍGUEZ-TAPIA, SERGIO; TODOROV APOSTOLOV, YORDAN. Reseña de García, Dolores; Hernández, Héctor; Sinner, Carsten. 2019. <i>Clases y categorías en la semántica del español y sus interfaces</i> . Berlin (Alemania): De Gruyter.....	319
CAMACHO NIÑO, JESÚS. Reseña de Contreras, Narciso M. 2020. <i>Estudios sobre el léxico del español. Diccionarios, variedades y enseñanza</i> . Jaén (España): UJA Editorial	324
OLIVA SANZ, CARMEN. Reseña de Guerrero, Gloria; Pérez, Manuel Fernando. 2020. <i>Terminología, Neología y Traducción</i> . Granada (España): Comares	335
GONZÁLEZ ARGÜELLO, VICENTA. Reseña de Lacorte, Manel; Reyes-Torres, Agustín. 2021. <i>Didáctica de la lengua como 2/L en el Siglo XXI</i> . Madrid: Arco Libros	343
MOLINA GARCÍA, ÁLVARO. Reseña de Iberoromania. <i>Contacto y migración. Desafíos metodológicos en la sociolingüística hispánica actual 91</i> (2020)	351
IVANOVA, OLGA. Reseña de Bañón, Antonio Miguel. 2018. <i>Discurso y salud. Análisis de un debate social</i> . Pamplona: Editorial Universidad de Navarra.....	359
PUJOL BERCHE, MERCÈ. Reseña de Kabatek, Johannes. 2018. <i>Lingüística coseriana, lingüística histórica, tradiciones discursivas</i> . Madrid / Frankfurt: Iberoamericana / Vervuert.....	365

La metodología lingüística en el siglo XXI
Andrea Ariño-Bizarro – Natalia López-Cortés
Universidad de Zaragoza
aribiz@unizar.es – natlop@unizar.es

La motivación que ha propiciado la elaboración de esta publicación surgió en 2019 tras la celebración de las V Jornadas Internacionales del Grupo de investigación de referencia Psylex (H11_20R) de la Universidad de Zaragoza. Durante los tres días en los que tuvo lugar el encuentro, un total de 13 expertos, algunos de ellos autores de este monográfico, ofrecieron diversas ponencias en las que compartieron sus experiencias personales con distintas metodologías y técnicas experimentales del ámbito lingüístico. El alto número de participantes y asistentes que acudieron entusiasmados a estas jornadas, así como la repercusión posterior que han tenido las ponencias grabadas y accesibles en línea en el repositorio *Zaragoza Lingüística a la Carta* (Ariño-Bizarro y López-Cortés 2022; PIIDUZ_3_345, coord. Ibarretxe-Antuñano), nos impulsaron a emprender este proyecto.

Si por algo se caracteriza la lingüística del siglo XXI, frente a la de otras épocas, es por la multitud de metodologías que han nacido para dar cabida a datos de distinta naturaleza que se emplean para falsar las hipótesis de la lingüística teórica. En los últimos años, la metodología lingüística ha experimentado un cambio importante no solo porque se han ampliado las posibilidades de realizar experimentos con instrumentos metodológicos novedosos, sino también porque se han ampliado los campos de investigación lingüística. A todo ello se suma la imperiosa necesidad de aplicar nuevas estrategias metodológicas a los estudios considerados como humanísticos para, de una vez por todas, conseguir erradicar la disyuntiva entre *ser de ciencias o de letras*. Gracias a este cambio de perspectiva, se ha conseguido difuminar progresivamente esta pregunta, tan recurrente años atrás, para dar paso a una única forma de ser: *ser científicos y científicas* (en este caso particular, interesados en el lenguaje y las lenguas) que trabajan con datos empíricos a través de diversas metodologías y herramientas de análisis.

Por esta razón, para ser un lingüista del siglo XXI, ya no es suficiente únicamente con la abstracción a la hora de elaborar teorías sobre cualquiera de los niveles de estudio de las lenguas, sino que todas estas hipótesis de partida han de ser probadas a través de la recogida, el estudio y el análisis de los datos reales ofrecidos por los hablantes. Y para que todo ello sea posible, es necesario que este proceso sea sostenido por un sólido armazón que garantice su validez: es aquí donde entran en juego las metodologías y técnicas experimentales que dan título a este monográfico.

Pese a que esta revolución metodológica es innegable, no se ha visto acompañada por la aparición de las correspondientes obras explicativas, al menos en lo que al español se refiere —véanse Litosseliti (2010), Sharma y Podesva (2014), para ejemplos en lengua inglesa—. Mostrar esta variedad de técnicas y estrategias metodológicas es, por tanto, el objetivo principal del presente monográfico. Aunque trabajos previos han dado cuenta de este creciente interés en el estudio científico del lenguaje, pocos de ellos están centrados en explicar cómo las técnicas e instrumentos metodológicos han de aplicarse a la hora de trabajar con datos empíricos en lingüística. Estos estudios, por tanto, se encargan de hablar en general de: en qué consiste el método científico y cómo se aplica en el estudio lingüístico. Por ejemplo, obras como la de Morales (1994), Jiménez (2007) y Reguera (2008) se centran en exponer por qué la lingüística es una ciencia (cuestión que ya ha sido

ampliamente probada) y en establecer pautas dirigidas para investigadores, normalmente noveles, sobre cómo establecer hipótesis, cómo controlar variables, y cómo presentar el análisis de datos, etc... Existen otros trabajos monográficos sobre metodologías utilizadas particularmente en ciertas áreas de la lingüística, como, por ejemplo, el trabajo de Hernández-Campoy y Almeida (2009) relativo al campo de la sociolingüística; el de Belinchón et al. (1994) relativo al de la psicolingüística o el de Rojo (2018), al de la traducción. E incluso, recientemente, se han preparado monográficos que recopilan distintas aproximaciones a la disciplina lingüística, ya no solo desde una perspectiva teórica, sino también aplicada y experimental (Gallego y Gutiérrez 2021).

Todo esto muestra que hay un claro interés por lo metodológico, pero siempre supeditado a una formación previa y desde una perspectiva lingüística determinada. La consecuencia última es que el lector que accede a estas obras se interesa por dicho material no tanto por la metodología en sí misma, sino más bien para conocer qué pasos se siguen en el método científico y acceder a un listado de posibles metodologías que se utilizan dentro de su rama de estudio.

Sin embargo, el interés actual va mucho más allá, puesto que, a la hora de recoger datos lingüísticos, un investigador ha de saber qué tipo de resultados quiere obtener (lo que determinará qué tipo de técnica experimental es más adecuada) así como establecer con posterioridad una correcta interpretación de los mismos (lo que se vincula con el alcance de estos datos en relación con las hipótesis de estudio planteadas). Puesto que la investigación experimental en lingüística es hasta cierto punto reciente (sobre todo, si la comparamos con otros ámbitos de la ciencia cognitiva, como la psicología o la neurociencia), existe la posibilidad de que los lingüistas no conozcamos todo el abanico de técnicas y metodologías que tenemos a nuestra disposición. O bien puede que sí que sepamos de su existencia, pero no seamos conscientes de qué tipo de información pueden ofrecernos y si esta es o no relevante para nuestras investigaciones. Este monográfico quiere servir como “catálogo” que proporcione una recopilación de algunas de las metodologías y técnicas experimentales que forman parte del estudio lingüístico, para así enumerar sus fortalezas (y también sus debilidades) y, sobre todo, para mostrar sus posibles aplicaciones a la investigación sobre el lenguaje y las lenguas.

Lo más novedoso de este monográfico es que esta explicación no está relacionada directamente con una teoría u área lingüística concreta; sino con un interés investigador particular: conocer, estudiar y descifrar en qué consiste una de las capacidades que nos distingue como especie, el lenguaje.

Teniendo en cuenta todo lo anterior, el propósito de *Metodologías lingüísticas: de los datos empíricos a la teoría del lenguaje* es ofrecer una explicación general y comprensiva sobre metodologías lingüísticas en español que muestre cómo los datos empíricos y los métodos experimentales, que tanto han avanzado en la actualidad, ayudan a construir explicaciones sólidas de cómo funcionan el lenguaje y las lenguas. En este sentido, este número monográfico reúne a diferentes especialistas para que respondan a interrogantes como para qué sirven, cómo se utilizan y por qué son relevantes diferentes instrumentos metodológicos que se utilizan en el área de la lingüística: desde las herramientas que recogen datos fisiológicos, como el *eye-tracker* o los potenciales evocados, hasta la metodología que fundamenta el análisis de documentos o el diseño de cuestionarios y entrevistas, pasando por el análisis de *big data* o la elaboración de corpus, entre otros.

En general, este monográfico está dirigido a todos aquellos investigadores, de cualquier especialidad, que sientan interés por la investigación lingüística y, por consiguiente, por el uso de metodologías de recogida y análisis de datos empíricos. Las técnicas metodológicas que se presentan, por tanto, interesan tanto a lingüistas que quieran ampliar su conocimiento sobre metodologías empíricas; como a aquellos especialistas de otras áreas de investigación, como la psicología, la sociología, la antropología o la ciencia cognitiva, entre otras.

El monográfico está dividido en dos partes diferenciadas: el BLOQUE I, titulado *Metodologías lingüísticas de la tradición*, recoge artículos dedicados a metodologías lingüísticas de mayor recorrido en el tiempo, por ser técnicas que llevan siendo utilizadas desde los inicios del estudio lingüístico; en el BLOQUE II, titulado *Metodologías lingüísticas de la contemporaneidad*, se recogen contribuciones dedicadas a técnicas experimentales de aplicación más reciente en el estudio de la lingüística. Con el objetivo de presentar una recopilación de esta información metodológica de forma accesible, coherente y comprensiva, todos los artículos dan respuestas, explícita o implícitamente, a las mismas preguntas: (i) ¿Para qué sirve la metodología lingüística en cuestión? (ii) ¿Cómo se trabaja con dicha metodología lingüística para la obtención o análisis de datos lingüísticos? y (iii) ¿Por qué es relevante este instrumento metodológico?

El BLOQUE I comienza con un trabajo, firmado por José M. García-Miguel, dedicado a la **lingüística de corpus**, metodología que aúna un conjunto de técnicas relacionadas con la compilación y explotación de corpus en los estudios lingüísticos. En este artículo se expone qué es un corpus lingüístico y qué relación tiene con otro tipo de datos, así como por qué es necesario anotarlo y cómo se establece y desarrolla el proceso de anotación.

A continuación, Gotzon Aurrekoetxea habla sobre el estudio de la variación. Así pues, este autor señala que, a día de hoy, los estudiosos en este campo del saber se han dotado de instrumentos y técnicas analítico-cuantitativas muy potentes, pero su objetivo sigue siendo el mismo: conocer cada vez mejor la variación lingüística. Entre estas herramientas se encuentra la técnica de la **dialectometría**, a la que se dedica el segundo artículo de este monográfico, y que permite: la delimitación cuantitativa de las fronteras dialectales, la clasificación jerárquica de los dialectos, el análisis de la distancia lingüística entre localidades o la detección de características lingüísticas diversas en distintas áreas dialectales.

Otra herramienta esencial para el estudio de la variación y que ha tenido, además, un gran peso en la tradición lingüística, ha sido el **análisis de documentos históricos**. María Teresa Moret Oliver y Javier Giralt Latorre explican, en el tercer artículo de esta obra, qué información lingüística relevante nos pueden aportar los textos no literarios de los registros históricos de una región geográfica concreta. Con este objetivo, ofrecen claves para un buen uso de este tipo de fuentes documentales. Se basan, para ello, en el caso concreto del catalán en Aragón.

El BLOQUE I se cierra con el trabajo de Iraide Ibarretxe-Antuñano sobre **trabajo de campo**. En este artículo se ofrece una definición de esta metodología desde una perspectiva general y se enumeran, tanto los requisitos necesarios para llevar a cabo la interacción entre el investigador y la comunidad lingüística que se quiera investigar, como los pasos necesarios para llevar a cabo una investigación de este tipo.

José Manuel Igoa abre el BLOQUE II, dedicado a las metodologías más recientemente incluidas en el ámbito de la lingüística, con un repaso de las **tareas conductuales**. Al

aplicar esta metodología, el autor señala que los investigadores pueden estudiar el comportamiento de los participantes a la hora de ofrecer respuestas vinculadas con estímulos lingüísticos, lo que proporciona información muy valiosa sobre el procesamiento del lenguaje.

El segundo artículo de este bloque, firmado por Victoria Marrero, explora qué pueden aportar las **técnicas de neuroimagen** a los estudios lingüísticos. Así, en este trabajo se presentan tanto las técnicas de base electromagnética como las de base metabólica. A través de todas ellas, la autora nos cuenta cómo, en la actualidad, a través de métodos y técnicas no invasivas se puede observar la respuesta neuronal vinculada a la actividad lingüística.

A continuación, Juan Haro se centra en una técnica de neuroimagen concreta, ampliamente utilizada en la actualidad: los **potenciales evocados**, esto es, las respuestas eléctricas del sistema nervioso que se desencadenan tras la presentación de un estímulo, en este caso, lingüístico. En este trabajo se explica, por tanto, cómo se realiza el registro y análisis de los datos obtenidos a partir de esta técnica y se presentan los distintos componentes que, como lingüistas, debemos tener en cuenta para estudiar el procesamiento del lenguaje.

Por su parte, Esther Álvarez dedica su contribución a explicar en qué consiste la metodología del **seguimiento ocular** (*eye-tracking*). Esta técnica asume que las personas miramos aquello que estamos pensando en ese momento, por lo que permite investigar, por ejemplo, a qué palabras, construcciones y elementos prestamos atención cuando leemos un texto; lo que posibilita estudiar cómo se procesa el lenguaje en el cerebro en tiempo real. Así, en este artículo, se expone cómo ciertos procesos cognitivos involucrados en el procesamiento se pueden estudiar a través de los patrones de movimientos oculares, tanto en tareas de lectura como de comprensión y de producción.

El siguiente artículo del monográfico, firmado por Javier Valenzuela, está dedicado al **análisis de big data**, también conocido como *macrodatos*, que estudia lingüísticamente toda la información que generamos, como hablantes, al utilizar los dispositivos digitales que tenemos a nuestra disposición diariamente. En este sentido, su trabajo discute algunas de las oportunidades así como de los desafíos inherentes que supone la aplicación del concepto de *big data* a los estudios del lenguaje y ejemplifica una de sus posibles aplicaciones al estudio de la multimodalidad.

Para finalizar, Nadège Juan e Yvette Bürki concluyen este bloque y, por tanto, esta obra colaborativa hablando sobre las **entrevistas mediadas**. En concreto, su contribución se centra en la recolección de datos sociolingüísticos a través de servicios de mensajería instantánea y supone, por tanto, una revisión y actualización de la noción de entrevista, que se describe tradicionalmente como un intercambio de información cara a cara.

En definitiva, este monográfico se establece como una propuesta dedicada a la exposición sistemática de distintas técnicas experimentales utilizadas en la actualidad. Con este fin, esperamos, por tanto, que los futuros lectores encuentren en este monográfico un material útil en el que puedan descubrir y conocer las posibles herramientas de las que se pueden valer para llevar a cabo sus investigaciones científicas.

Finalmente, nos gustaría agradecer a todos los autores que han participado en este monográfico, primero, su entusiasmo al aceptar la invitación y, segundo, su trabajo impecable, su disponibilidad y toda su tenacidad durante el desarrollo del mismo.

También queremos darle las gracias tanto al Comité Editorial de la revista *Estudios de Lingüística del Español* por confiar en nosotras y por ver futuro en este proyecto, como al Comité Científico, cuyas minuciosas revisiones han enriquecido todas y cada una de las contribuciones que conforman este monográfico. Además, no podemos dejar de mencionar toda la ayuda y el soporte recibido durante la ejecución de este proyecto, por parte del Grupo de Investigación de referencia PSYLEX (H11_20R) de la Universidad de Zaragoza y del proyecto de investigación CONESSO (FFI2017-82460-P) del Ministerio de Ciencia e Innovación.

Referencias

- Ariño-Bizarro, Andrea; López-Cortés, Natalia. 2022. Repositorios multimedia como herramienta de innovación pedagógica: El caso de Zaragoza Lingüística a la Carta. *Didáctica de la lengua en nuevos entornos de aprendizaje*. Barcelona: Graó.
- Belinchón, Mercedes; Igoa, José Manuel; Rivière, Ángel. 1994. *Psicología del lenguaje: investigación y teoría*. Madrid: Trotta.
- Gallego, Ángel, J.; Gutiérrez, Edita, coord. 2021. La investigación lingüística en el s. XXI: aproximaciones teóricas y sus aplicaciones. *Revista Española de Lingüística* 51.2. Disponible en: <http://revista.sel.edu.es/index.php/revista/issue/view/123>.
- Hernández Campoy, José Manuel; Almeida, Manuel. 2005. *Metodología de la investigación sociolingüística*. Granada: Editorial Comares.
- Jiménez Ruiz, Juan Luis. 2007. *Metodología de la investigación lingüística*. Alicante: Publicaciones de la Universidad de Alicante.
- Litosseliti, Lia. 2010. *Research methods in Linguistics*. Londres: Continuum.
- López Morales, Humberto. 1994. *Métodos de investigación lingüística*. Salamanca: Ediciones del Colegio de España
- Reguera, Alejandra. 2008. *Metodología de la investigación lingüística: prácticas de escritura*. Córdoba (Argentina): Editorial Brujas.
- Rojo, Ana. 2018. *La investigación en traducción. Una revisión metodológica de la disciplina*. Barcelona: Anthropos.
- Sharma, Devyani; Podesva, Robert J. 2014. *Research methods in Linguistics*. Reino Unido: Cambridge University Press.

Lingüística de corpus: de los datos textuales a la teoría lingüística

José M. García-Miguel
Universidade de Vigo
gallego@uvigo.es

Resumen

Este artículo es una presentación general de la lingüística de corpus en el que se expone qué es un corpus lingüístico, qué relación tiene con otros tipos de datos, por qué es necesario anotarlo y cómo es el proceso de anotación. También se pasa revista a algunas de las tareas más comunes en la investigación lingüística basada en corpus, tales como la obtención de listados de frecuencias, la exploración de concordancias o la búsqueda de coapariciones (colocaciones) y otros tipos de información contextual. A lo largo del texto se intenta mostrar la relevancia de este tipo de datos para la teoría lingüística, en particular, para los modelos basados en el uso, como los cognitivos y funcionales.

Palabras clave: corpus, anotación, datos, frecuencia, coapariciones, teoría lingüística.

Abstract

In this paper a general presentation of Corpus Linguistics is provided by explaining what a linguistic corpus is, how it is related to other types of data, why it is necessary to annotate it, and what the annotation process is like. Some of the more common tasks in corpus-based linguistic research are also reviewed, such as obtaining frequency lists, exploring concordances, or finding co-occurrences (collocations) and other types of contextual information. Throughout the text, an attempt is made to show the relevance of this type of data for linguistic theory, in particular for use-based models, such as the cognitive and functional ones.

Keywords: Corpus, annotation, data, frequency, co-occurrences, linguistic theory.

1. Introducción: ¿qué es la lingüística de corpus?

La lingüística de corpus es un conjunto de metodologías relacionadas con la compilación y explotación de corpus en los estudios lingüísticos tanto teóricos como aplicados. La característica distintiva de esta rama de la lingüística es pues la utilización de corpus textuales como fuente primaria de datos. Un corpus puede definirse (por ejemplo, Sinclair 1991: 171; McEnery et al. 2006: 5; Gries 2009: 7; Rojo 2021: 1) como una colección de textos orales o escritos producidos en un contexto comunicativo natural, representativos de una lengua o variedad de lengua, almacenados en soporte informático y destinados al análisis lingüístico.

Los textos que componen un corpus (por ejemplo, una obra literaria, una noticia de periódico o una conversación entre amigos) fueron producidos con la intención de comunicar algo en cierto contexto y no generados más o menos artificialmente para investigar o ilustrar algún fenómeno lingüístico. Por tanto, los lingüistas de corpus basan sus descripciones y teorías en el uso real del lenguaje por parte de los hablantes, y en eso

se contraponen a otros métodos de obtención de datos tales como la introspección, las encuestas o la experimentación. En apartados posteriores volveremos sobre este punto y sus repercusiones teóricas.

Desde hace mucho tiempo se han hecho recopilaciones de textos con diferentes propósitos, incluido el análisis lingüístico; pero para llegar al estado actual de la disciplina fue crucial el desarrollo y difusión de la informática con su capacidad para almacenar y procesar datos lingüísticos. Por tanto, actualmente un corpus no es una simple colección de textos, sino un conjunto de textos digitalizados, que se pueden procesar mecánicamente. Toda la lingüística de corpus actual entra dentro de lo que Leech (1992: 106) prefiere etiquetar como “computer corpus linguistics”.

Sin embargo, la lingüística de corpus, tal como la entendemos actualmente, ya tiene una tradición de más de medio siglo, que podemos remontar hasta la aparición en 1964 del primer corpus de la era moderna, el corpus Brown, que está conformado por 500 fragmentos de textos de inglés americano de un tamaño de aproximadamente 2000 palabras cada uno, hasta totalizar un millón de palabras. Tras estos primeros corpus, fueron apareciendo, no solo para el inglés sino también para otras lenguas europeas, corpus de referencia que pretendían ser muestras equilibradas y representativas de toda la lengua (tanto en sincronía como en diacronía). Un ejemplo prototípico de corpus es el *British National Corpus* (BNC), que contiene muestras de inglés británico hablado (10 %) y escrito (90 %) que suman un total de 100 millones de palabras.

En el caso del español, tenemos a nuestra disposición los corpus de la RAE (CREA, CORDE, CORPES) o la versión histórica del *Corpus del español* [CdE-hist] de M. Davies. De ellos, CORDE y CdE-hist son corpus diacrónicos que recogen textos desde los orígenes del español hasta el siglo XX, mientras que CREA y CORPES son corpus del español contemporáneo, que recogen en distintas proporciones textos tanto orales como escritos procedentes de España y América, de diferentes géneros y ámbitos temáticos. Cada uno de ellos contiene entre 100 y 400 millones de palabras.

Actualmente, la disponibilidad de textos que nos proporciona internet ha facilitado la creación de corpus de mucho mayor tamaño por el procedimiento de descargar textos de manera bastante indiscriminada a costa de sacrificar en mayor o menor medida el equilibrio, la representatividad y la diversidad de la muestra (para saber más sobre *big data*, véase Valenzuela 2022). La plataforma de *Corpus del español* dispone de un corpus (CdE-web) de 2 mil millones de palabras y otro de noticias (CdE-NOW) con más de 5 mil millones. En la plataforma *Sketch Engine* se pueden consultar corpus del español (esTenTen18) y de otras lenguas de un tamaño cercano a los 20 mil millones de palabras, llegando al doble en el caso del inglés (enTenTen20).

Existen muchos corpus de diferentes tipos, tamaños y composiciones (puede verse una tipología detallada en Torruella 2017 y un panorama general de corpus del español en Rojo 2016). Junto a los corpus de referencia representativos de una lengua, como el CREA o el CORPES, y los corpus diacrónicos como el CORDE, existen corpus que se centran en alguna variedad en particular. Por ejemplo, hay corpus de lenguaje especializado (textos jurídicos, informáticos, económicos, etc.), como el *Corpus Técnico do Galego*, o el *Corpus Tècnic del IULA* (multilingüe). Los corpus multilingües, como este último, pueden tener textos obtenidos por procedimientos similares que faciliten la comparación entre lenguas (*corpus comparables*, como la familia de corpus TenTen), o consistir en traducciones de una lengua a otra(s) convenientemente alineadas (*corpus*

paralelos, como CLUVI). También hay corpus de textos producidos por personas de determinado rango de edad, como los diferentes corpus de habla infantil contenidos en CHILDES o como COLA, que contiene lenguaje de adolescentes. Así mismo hay corpus de aprendices, como CAES o CORELE, que contienen textos producidos por quienes están aprendiendo una lengua y no por hablantes nativos. Y también existen otros muchos corpus compilados con propósitos y contenidos específicos. Mención especial merecen los corpus orales, como CORLEC, ESLORA, Val.Es.Co o los corpus del proyecto PRESEEA, dadas las dificultades de la recogida de datos orales tanto en la fase de grabación como en la de transcripción. Mayores aún son las dificultades para compilar corpus de lenguas de signos, aunque, en los últimos años, estamos viendo aparecer corpus de varias lenguas signadas como Auslan, BSL, NGT, DGS, o más recientemente CorLSE. Cada corpus nos proporcionará datos para estudiar en profundidad cierta(s) variedad(es) de lengua(s) y no otras; por lo que para quien utilice esta metodología siempre será importante seleccionar un corpus apropiado para los objetivos de su investigación.

A los corpus existentes y disponibles podemos añadir la posibilidad de que un lingüista individual pueda construir su propio corpus para atender a los objetivos de su investigación. En cualquier ordenador podemos disponer de herramientas, muchas de ellas gratuitas, que nos sirven de ayuda para recoger textos para compilar un corpus, añadirle automáticamente anotación lingüística y examinar los textos con herramientas típicas de la lingüística de corpus: concordancias, agrupaciones, listas de palabras, palabras clave, etc. Además, casi todo ordenador personal suele tener instalados una serie de programas de uso habitual que pueden ser muy útiles para trabajar con corpus, tales como editores de texto, hojas de cálculo, gestores de bases de datos, programas estadísticos, etc. Un plus bastante recomendable para quien se anime a trabajar con su propio corpus personal es el de tener conocimientos básicos de programación; pero, incluso con escasos o nulos conocimientos de programación, cualquier lingüista puede actualmente hacer lingüística de corpus con su ordenador personal, bien sea manejando sus propios datos o bien consultando corpus existentes elaborados por otras personas o por grandes equipos de trabajo.

En las siguientes secciones se ofrecerá un panorama general, inevitablemente algo superficial, de la metodología lingüística basada en corpus tomando como principal punto de vista el de un usuario lingüista. En la sección 2 se sitúan los datos tomados de corpus en relación con otras fuentes de datos lingüísticos. En la sección 3 se observa el proceso de compilación de corpus desde la selección de textos hasta su anotación y puesta a disposición de los usuarios. En la sección 4 se detallan algunas de las operaciones más básicas que un lingüista puede hacer con un corpus: buscar y obtener muestras de uso contextualizado de unidades lingüísticas, contar frecuencias de uso y documentar y computar coapariciones en el contexto inmediato. En la sección 5 se exponen brevemente las conexiones de la lingüística de corpus con algunas tendencias teóricas contemporáneas.

2. La lingüística de corpus y las fuentes de datos para hacer lingüística

Como en toda ciencia, las descripciones y teorías lingüísticas deben estar basadas en datos empíricos. Con carácter general, en lingüística podemos distinguir entre datos basados en la observación del uso del lenguaje por parte de los hablantes y datos basados en los

juicios y reacciones de los hablantes sobre la lengua, tales como los proporcionados por las encuestas o por los métodos experimentales. En este último grupo podría incluirse también la introspección, los juicios que el propio lingüista hace sobre su lengua. Fillmore (1992: 35) traza una conocida caricatura de dos especies de lingüistas según el tipo de datos que utilizan: por un lado, el “lingüista de sillón” que elabora teorías lingüísticas fiándolo todo a la introspección y sin apenas otros datos sobre los que sustentarlas y, por otro, el “lingüista de corpus” que acumula datos y más datos sin llegar a decir nada interesante sobre ellos. El primero ve cuestionada la validez empírica de sus propuestas, el segundo de los dos ve cuestionado el interés teórico de sus observaciones. Es posible que la polarización entre estos dos tipos de lingüistas se tome a veces demasiado en serio (Gries 2010b: 331), pero la descripción humorística de Fillmore nos recuerda la dialéctica entre reflexión teórica y recopilación de datos, que debe darse en toda ciencia y que apunta al mismo tiempo al hecho de que hay múltiples maneras de obtener datos para hacer lingüística. Más en serio, Fillmore observa que ningún corpus, por grande que sea, contiene todo lo que un lingüista puede llegar a necesitar; pero que, en todo corpus, aunque sea pequeño, podemos encontrar hechos que no se encuentran de otra manera. Su conclusión es que los dos tipos de lingüistas se necesitan mutuamente, o mejor aún, que los dos tipos de lingüistas deberían coexistir, siempre que sea posible, dentro del mismo cuerpo (Fillmore 1992: 35). La recomendación general es, pues, que la reflexión teórica vaya acompañada de datos (y viceversa), y utilizar diferentes metodologías de obtención de datos, muchas veces combinándolas o contrastándolas, y, en todo caso, siempre recurrir a la metodología que mejor se adapte a los objetivos de nuestra investigación.

La lingüística de corpus se caracteriza porque toma como fuente primaria de datos lo que los hablantes efectivamente dicen y escriben cuando se comunican por medio del lenguaje. Gilquin y Gries (2009: 5) ordenan los tipos de datos lingüísticos en una escala de más natural a menos natural, situando en un extremo los corpus de textos escritos y en el otro extremo los métodos experimentales en los que se pide a los participantes que hagan cosas con el lenguaje que normalmente no hacen (tales como reaccionar en condiciones de laboratorio ante ciertos estímulos lingüísticos o visuales). Frente a otros métodos de obtención de datos, la lingüística de corpus trata con datos naturales en cuya producción no ha influido el analista. Esto es, los textos que componen un corpus (sean obras literarias o científicas, artículos de prensa, páginas web, folletos, manuales de instrucciones, cartas...) se produjeron en su momento con un propósito que no era el de formar parte de un corpus. Con todo, debido a lo que se conoce desde Labov (1972: 171) como la *paradoja del observador*, algunos contextos comunicativos y algunas variedades lingüísticas son difíciles de observar sin influir en la producción. Por eso algunos corpus incluyen también transcripciones de conversaciones en las que participa el analista (observador participante), como ocurre en entrevistas sociolingüísticas, o en producciones discursivas que se obtienen como reacción a ciertos estímulos. En cualquier caso, para que podamos hablar de un corpus este debe contener textos (no necesariamente completos), puesto que un conjunto de palabras u oraciones aisladas, incluso aunque estas se hayan documentado en contextos naturales reales, no constituyen un corpus.

Por todo ello, la lingüística de corpus puede verse como una aproximación empírica cuyo punto de partida son los datos de uso del lenguaje tal como se realiza en los textos. Pero en tanto que aproximación que ha desarrollado sus propios métodos, la lingüística de corpus es muy diferente del análisis del discurso, pues su objetivo no es el análisis e interpretación de instancias individuales de uso, sino encontrar en la colección de textos

que constituye el corpus patrones regulares de uso que nos ayuden a entender la estructura y funcionamiento de un sistema lingüístico o a entender prácticas sociales, actitudes e ideologías que se reflejan en los usos lingüísticos. Tognini-Bonelli (2001) resume, como se muestra en la Tabla 1, las diferencias entre análisis textual y lingüística de corpus.

Texto	Corpus
Se lee entero	Se lee fragmentado
Se lee horizontalmente	Se lee verticalmente
Se lee el contenido	Se buscan patrones formales
Se lee como evento único	Se buscan eventos repetidos
Se lee como acto individual	Se estudia como muestra de una práctica social
Evento comunicativo coherente	No es un evento comunicativo coherente
Instancia de actuación individual (<i>parole</i>)	Ayuda a entender el sistema lingüístico (<i>langue</i>)

Tabla 1. Análisis de texto y análisis de corpus (adaptado de Tognini-Bonelli 2001: 3)

Los métodos de la lingüística de corpus también presentan diferencias con los utilizados para documentación de lenguas y variedades lingüísticas poco descritas. Para la mayoría de las lenguas es imposible obtener datos por el simple procedimiento de descargar textos de internet. En las etapas iniciales de descripción, necesitamos recoger léxico, patrones gramaticales y textos normalmente mediante procesos de elicitación con hablantes nativos. La recogida de textos puede llegar a constituir un corpus representativo útil para los propósitos de la descripción. Lo esperable y deseable es que cada vez tengamos más datos y más corpus sobre más lenguas. Sin embargo, la lingüística de corpus no se ha desarrollado históricamente como método para documentar lenguas de las que no teníamos datos, sino como conjunto de herramientas para el análisis de lenguas, como el inglés o el español, para las que existe una larga tradición descriptiva apoyada en una larga tradición escrita y, por tanto, en la disponibilidad de textos.

Lo que ha cambiado frente a la lingüística de hace un siglo es que la tecnología actual nos permite manejar en poco tiempo una cantidad de datos antes impensable. Y eso es lo que nos ha traído la informática. El desarrollo de los corpus como herramienta para los lingüistas se ha comparado con la invención del telescopio, que permitió observar cosas que no se habían visto nunca antes (Stubbs 1996: 231 y, tras él, muchos otros). El telescopio aproxima el objeto de estudio, permite una observación más precisa y, con ello, la formulación de teorías más coherentes y comprensivas sobre el universo. La revolución informática moderna permite el rápido procesamiento de ingentes cantidades de datos, lo que ha revolucionado todas las ciencias, incluida la lingüística. El físico Freeman Dyson (1997: 50-51) explica que en la historia de la ciencia pueden producirse revoluciones conceptuales [*'concept-driven revolutions'*], las cuales explican cosas viejas de maneras nuevas, y revoluciones instrumentales [*'tool-driven revolutions'*], las cuales descubren cosas nuevas que hay que explicar, y añade que las revoluciones instrumentales son más frecuentes que las conceptuales. Pone como ejemplos el impacto que tuvieron en su tiempo el telescopio, luego el microscopio, y más recientemente los computadores. La lingüística de corpus forma parte de una revolución instrumental en la que la informática pone a nuestra disposición muchos datos nuevos que debemos explicar.

Una propiedad destacable de la lingüística de corpus es la exhaustividad, entendida como la obligación de intentar un examen de todos los casos contenidos en el corpus. El análisis y las interpretaciones teóricas deben dar cuenta de todos los datos, sin hacer una selección previa que descarte aquellos casos que decidimos ignorar por irrelevantes. Este principio

de exhaustividad en relación con los datos disponibles es un destacable punto fuerte de la lingüística de corpus (Leech 1992: 112). Cualquier corpus nos va a presentar ejemplos abundantes de fenómenos muy comunes, evitando que nuestras descripciones se basen en fenómenos llamativos que atraen nuestra atención y minusvaloren lo más común. Pero el principio de exhaustividad lleva a que cualquier corpus nos pueda presentar muchos casos de fenómenos ‘raros’, poco frecuentes, y nos obligue a preguntarnos si nuestras categorías analíticas valen para todos los casos posibles.

3. De los textos a los corpus anotados

El lingüista que quiere utilizar en sus investigaciones datos de corpus debe tener en cuenta que tanto la compilación como la utilización o explotación de corpus conllevan en todas sus fases una serie de problemas prácticos que pueden o suelen tener carga teórica: ¿cómo seleccionamos una muestra de textos que sea representativa y equilibrada?, ¿cómo digitalizamos los textos escritos?, ¿cómo transcribimos los textos orales y signados?, ¿cómo reconstruimos la relación entre el texto-producido y el contexto dinámico en que se produjo?, ¿cómo identificamos y clasificamos las unidades básicas (palabras, oraciones...)? Analizaremos brevemente en esta sección algunos de estos problemas desde el punto de vista del lingüista usuario de esta metodología, teniendo en cuenta también el punto de vista de los creadores de corpus.

El primer problema al que nos enfrentamos es la selección de los textos que constituyen el corpus o, para un usuario final, la selección de un corpus u otro. Por grande que sea, ningún corpus contiene todos los discursos producidos en una lengua y solo en casos muy específicos podría llegar a contener todos los textos existentes de cierta variedad lingüística. Los corpus consisten en muestras que pretenden ser representativas del conjunto de la lengua y, por eso, siempre que se diseña un corpus deben considerarse con atención los criterios de selección (véase Biber 1993; Toruella y Llisterri 1999). Sin embargo, como muy raramente conocemos el tamaño y composición exacta de la población total de la que se obtiene la muestra (el conjunto total de textos orales y escritos producidos en una variedad lingüística) ni son igualmente accesibles todos los tipos de discurso, la representatividad y el equilibrio de un corpus es solo un ideal teórico prácticamente imposible de alcanzar (véase Stefanowitsch 2020: 29; Rojo 2021: 291-294). En cambio, una vez decidida la lengua o variedad de lengua, se intenta que la muestra sea diversificada, que contenga textos de diferentes géneros, temáticas, procedencias geográficas, etc. de manera que ningún tipo particular tenga un peso excesivo que pueda sesgar los resultados que obtengamos. Esa es también una de las razones por las que, en los primeros corpus, como el de Brown, se incluían fragmentos textuales no demasiado grandes y no obras completas. En los corpus modernos de gran tamaño y diversificación es más difícil que una obra o género particular tenga un peso excesivo. En cualquier caso, es muy importante que al extraer conclusiones de un corpus tengamos en cuenta de qué es representativo ese corpus y qué carencias o sesgos puede tener, antes de pretender generalizar a toda una lengua lo que quizá sea específico de esa muestra.

Por otro lado, aunque las diferentes secciones de un corpus no tengan tamaños proporcionales a los de la población total, lo que nos interesa habitualmente no son las frecuencias absolutas de un fenómeno, sino frecuencias relativas o normalizadas que nos

permitan comparar una sección con otra independientemente del tamaño de cada muestra. Si el corpus es grande podríamos obtener datos suficientes sobre el fenómeno de interés y observar a posteriori cómo se distribuyen en las diferentes secciones del corpus para determinar en qué variedades lingüísticas ocurren esos datos. Para el lingüista que hace investigación basada en corpus es crucial conocer cómo está hecho el corpus que utiliza y saber cuáles fueron los criterios de selección de textos y qué sesgos puede tener esa selección. De eso depende en buena medida la validez de las conclusiones que obtenga. Casi más importante aún que el equilibrio de la muestra es que los textos que componen el corpus estén catalogados con información sobre las circunstancias de su producción (género, medio, ámbito temático, fecha, lugar de origen...) o sobre las características de quienes los produjeron (sexo, edad, nivel sociocultural...). Estos metadatos son los que nos permitirán seleccionar subcorpus o realizar búsquedas con diferentes criterios y comparar unas variedades con otras.

Los textos que componen un corpus deben compilarse necesariamente en soporte informático para poder ser analizados con los métodos de la lingüística de corpus. La informatización es un problema relativamente complejo si queremos contar con muestras de discurso oral (que primero deben transcribirse) o con muestras de discurso escrito que hasta hace poco estaban mayoritariamente en soporte papel. Por el contrario, parece un problema trivial (aunque no tanto como parece a primera vista) si construimos un corpus recurriendo a las ingentes cantidades de textos que se producen actualmente en internet. Pero, sea cual sea el soporte original, debemos guardarlos en un formato que permita al usuario la consulta desde cualquier sistema informático utilizando las herramientas que proporciona la lingüística de corpus. En general, esto requiere registrarlos como texto plano sin ningún tipo de formato ni de elemento adicional. Un archivo de texto simple digitalizado no es ni más ni menos que una cadena de códigos que simbolizan caracteres (alfabéticos, numéricos, signos de puntuación y otros). Los textos escritos pueden codificarse directamente, pero hay que prestar atención al sistema de codificación, para evitar problemas con diacríticos y otros símbolos que van más allá del alfabeto latino básico. Lo más recomendable actualmente es el sistema Unicode, válido para cualquier sistema de escritura.

En este sentido, los textos orales, multimedia y signados pueden grabarse; pero, para poder ser utilizados en lingüística de corpus, deben transcribirse a texto escrito utilizando un sistema de transcripción apropiado. Dependiendo de los objetivos con los que se compila el corpus, podría ser suficiente con una transcripción ortográfica, que luego codificamos en un archivo de texto plano igual que cualquier otro texto escrito; pero muy probablemente también nos interese poder recuperar aspectos de la pronunciación, la prosodia y otros aspectos del discurso oral, tales como pausas, vacilaciones y solapamientos, que no se recogen en la escritura ortográfica convencional. Por su parte, en el caso de las lenguas de signos se añade el problema de que no existe un sistema estándar de signoescritura; por lo que, para poder explorar un corpus de una lengua de signos, lo habitual es acompañar las grabaciones en vídeo de transcripciones basadas en un sistema de glosas identificativas ('id-glosas') de cada seña (Johnston 2010; Pérez et al. 2019). Más allá de la simple transcripción o glosado de lenguaje oral o signado, uno de los grandes retos para la lingüística de corpus actual es la recopilación, codificación y explotación de corpus multimodales, consistentes en colecciones de grabaciones audiovisuales que sirven para estudiar la interacción entre las diferentes modalidades (voz, gestos, mirada, orientación corporal, proxémica, etc.) utilizadas simultáneamente

en la comunicación humana y la interacción entre ellas (Kipp et al. 2009; Knight y Adolphs 2020). Para su análisis necesitaremos varios niveles de transcripción y anotación alineados temporalmente con las grabaciones de audio y video. Herramientas como ELAN pueden ayudar en esa tarea.

Los textos escritos también suelen contener en su versión original mucha información que se pierde al digitalizarlos como texto plano: disposición espacial, tipografía, abreviaturas, gráficos, tablas, llamadas a notas, etc. Esa información puede ser relevante para la interpretación del texto y en el diseño de un corpus se debe determinar cuánta de esa información debe codificarse y con qué procedimiento. Lo mismo puede decirse de la estructura del texto, por ejemplo, la de un libro en capítulos y secciones, o la de una noticia de periódico en titular, entradilla y cuerpo del texto. Por tanto, la compilación de un corpus requiere, además de la codificación del texto como cadena de caracteres, la codificación de información externa sobre cada texto (los metadatos) o información interna sobre propiedades que acompañan al texto (paralenguaje, estructura interna, etc.). La mayoría de los corpus existentes actualmente contienen anotaciones de alguno de los siguientes tipos:

- Información que cataloga los textos según las circunstancias de su producción (género, medio, ámbito temático, fecha, lugar de origen...) o según las características de quienes los produjeron (sexo, edad, nivel sociocultural...)
- Información sobre rasgos paralingüísticos que acompañan al texto, tales como el tipo y tamaño de fuente en los textos escritos o la prosodia en los textos orales transcritos. También puede incluirse información o referencias sobre elementos simbólicos multimodales, tales como gráficos que acompañan a textos escritos o gestos con textos orales.
- Información sobre la estructura del texto: capítulos, secciones, secuencias temáticas, movimientos conversacionales, ...

Esta codificación recibe a veces el nombre de *anotación*, pero normalmente se reserva este término o el de *etiquetación* para la codificación de propiedades lingüísticas relativamente abstractas que no pueden observarse directamente en el texto. Si un usuario quiere recuperar datos de un corpus consistente en archivos de texto podrá buscar con el software apropiado cadenas de caracteres que tal vez correspondan a (partes de) palabras y frases. Podemos obtener así rápidamente los datos buscados, pero difícilmente podemos encontrar de esta manera propiedades abstractas como, por ejemplo, una estructura sintáctica particular. Para ello sería necesario anotar el texto, esto es, añadir indicaciones explícitas de propiedades lingüísticas de los elementos del texto, tales como clase de palabra, lema, acepción de una palabra polisémica, o la estructura sintáctica o semántica de una oración. Cualquier nivel de análisis lingüístico es susceptible de anotación. Es posible, por tanto, que un corpus contenga anotación de diversos tipos:

- Fonológica: transcripción fonética, límites silábicos, rasgos prosódicos (acento, tono, entonación)
- Ortográfica: asociación de variantes ortográficas con formas normalizadas
- Morfológica: raíces, prefijos, sufijos...
- Léxico-gramatical: lematización, clase de palabra (N, V), rasgos morfosintácticos (singular, plural, tiempo verbal, ...)

- Sintáctica: constituyentes, relaciones de dependencia, funciones sintácticas (Suj, Obj), categorías sintácticas...
- Semántica: desambiguación léxica, campos semánticos/ontologías, papeles sintáctico-semánticos, clasificación de nombres propios, metáforas, ...
- Textual-discursiva: relaciones anafóricas, tema/rema, información dada/nueva, estructura del discurso, ...
- Pragmática: actos de habla, análisis de sentimientos y opiniones, roles discursivos, conocimiento contextual compartido, ...

De estos niveles de anotación, el más fácil de implementar y por ello el más extendido con mucha diferencia en los corpus disponibles para uso general es la anotación léxico-gramatical consistente en la lematización y etiquetado morfosintáctico de las palabras del corpus. Con la lematización adscribimos cada palabra ortográfica a una forma de diccionario o lema. En el etiquetado morfosintáctico se incluye, al menos, la clase de palabra y usualmente también las categorías flexivas. Por ejemplo, la palabra ortográfica *cuento* puede corresponder al singular del sustantivo *cuento* o a la primera persona singular del presente de indicativo del verbo *contar*.

Toda anotación implica siempre una interpretación del texto, por lo que requiere siempre un proceso de toma de decisiones informado teóricamente, aunque condicionado por limitaciones prácticas. Y esto vale para cualquier nivel de análisis, incluso los aparentemente más triviales. Por ejemplo, la tokenización o división del texto en unidades elementales susceptibles de recibir anotaciones ulteriores suele consistir en la división del texto en palabras ortográficas delimitadas por espacios y signos de puntuación. Aunque parece un proceso sencillo, inmediatamente nos encontramos con problemas prácticos, pues algunos signos de puntuación no separan palabras ortográficas: los guiones unen a veces palabras compuestas y los puntos podemos encontrarlos en el interior de siglas y abreviaturas. Más importante aún es el hecho de que el estatus teórico de la palabra como unidad básica del análisis lingüístico es discutible y que los límites entre palabras son indeterminados (Haspelmath 2011). En español, por ejemplo, podemos discutir el estatus como palabra de clíticos pronominales (que a veces se unen ortográficamente al verbo y otras no), pero también el estatus de artículos, preposiciones, conjunciones y otras formas dependientes.

El proceso de anotación en lingüística de corpus acarrea diversos problemas prácticos relacionados con cómo se realiza y con cómo se codifica. Otras facetas del proceso de anotación tienen que ver con el diseño de esquemas de anotación, desarrollo de estándares, evaluación y usos de las anotaciones, etc. (Ide y Pustejovsky 2017).

En cuanto al procedimiento, la anotación puede ser manual o automática. La anotación manual solo es factible con corpus pequeños y, aun así, es muy trabajosa y está sujeta a imprecisiones e inconsistencias. La anotación manual es compatible con cualquier punto de vista teórico y permite a uno decidir qué es lo que quiere anotar, según las necesidades de su investigación. Ahora bien, se puede facilitar la comparación entre corpus anotados si se utilizan los mismos esquemas de anotación, de ahí la conveniencia de desarrollar estándares de anotación. En cualquier caso, se requiere una definición previa de las categorías que se van a utilizar en la anotación y de los criterios de delimitación. Si los criterios están bien definidos deberían tener validez intersubjetiva, esto es, personas diferentes llegarían a clasificaciones similares de los datos. Conviene, además, utilizar algún índice estadístico para cuantificar el grado de acuerdo entre anotadores diferentes.

Con corpus grandes solo es factible la anotación automática. Existen programas suficientemente eficaces, muchos de ellos gratuitos, para ciertos tipos de anotación. En la lematización y el etiquetado morfosintáctico de las palabras del corpus se alcanzan porcentajes de acierto superiores al 95 %. Sin embargo, un error inferior al 5 % hace que en un corpus de 100 millones de palabras tengamos casi 5 millones de palabras anotadas erróneamente y que en una oración elegida al azar una palabra de cada 20 esté mal etiquetada. Para otros niveles de análisis, como la anotación sintáctica o la desambiguación de significados léxicos, el porcentaje de acierto de los programas existentes es menor y por eso son mucho más raros los corpus disponibles con ese tipo de anotación. Recursos que contienen análisis sintáctico y semántico de corpus, tales como ADESSE y BDS, AnCorA, SenSem o el IULA Spanish Treebank, han sido elaborados a mano en su totalidad o en un porcentaje significativo, aunque lógicamente han contado con asistencia informática.

El proceso de anotación automática está condicionado, en primer lugar, como en la anotación manual, por la elección del propio esquema de anotación (por ejemplo, qué sistema de clasificación de palabras utilizar). Luego el programa categoriza los elementos mediante cálculo de probabilidades. Por ejemplo, en la frase *El cuento de los tres cerditos*, se trataría de calcular la probabilidad de que *cuento* sea nombre o verbo, junto con la probabilidad de que un nombre o un verbo aparezca tras el artículo *el* o ante la preposición *de*. No obstante, estas probabilidades, a su vez, suelen estar basadas en “corpus de entrenamiento” que han sido anotados o revisados manualmente. La eficacia de un proceso automático de anotación dependerá pues del esquema de anotación y de la cantidad y calidad del corpus de entrenamiento; y de si ha sido posible una ulterior revisión manual de la anotación automática.

Para el usuario de un corpus ya anotado, sea manual o automáticamente, es importante tener en cuenta que toda anotación es el resultado de un proceso de toma de decisiones llevado a cabo por otros investigadores (Stefanowitsch 2020: 83). Cualquier usuario de un corpus debe tomar la precaución de no asumir acríticamente las anotaciones que encuentra en el corpus y debe conocer los criterios que se han utilizado en la anotación. Y viceversa, quienes anoten un corpus deben explicitar los criterios de anotación. Siendo así, la anotación siempre añade valor a un corpus, incluso si contiene categorizaciones discutibles de los datos.

Existen diferentes maneras de codificar la anotación de un corpus. Un procedimiento común hace un tiempo es utilizar etiquetas unidas a la unidad etiquetada mediante un separador predefinido, como un guion bajo, como, por ejemplo, en *El_ART cuento_NC de_PREP los_ART tres_NUM cerditos_NC*. Con cualquier software de exploración de corpus (incluso con cualquier editor de texto) utilizando comodines o patrones de búsqueda con expresiones regulares podríamos buscar fácilmente elementos del texto, categorías anotadas o una combinación de ambas.

Mucho más potente y más extendido es el lenguaje de marcado XML que nos permite codificar virtualmente cualquier cosa utilizando etiquetas entre ángulos (por ejemplo, encerrando una oración entre `<oración></oración>`) y pares atributo = valor acompañando a las etiquetas (por ejemplo, `gen=“masc”`). Este lenguaje puede servir para hacer anotaciones lingüísticas, como en el ejemplo (1), y también es el más utilizado para codificar los metadatos y la estructura de los documentos de un corpus (para lo cual, es recomendable seguir las recomendaciones de la “Text Encoding Initiative” [TEI]).

(1)
<oración id="1">
<palabra lema="el" cat="artículo" gen="masc" num="singular" > **El** </palabra>
<palabra lema="cuento" cat="nombre" gen="masc" num="singular" > **cuento**
</palabra>
<palabra lema="de" cat="preposición"> **de** </palabra>
<palabra lema="el" cat="artículo" gen="masc" num="plural" > **los** </palabra>
<palabra lema="3" cat="numeral" > **tres** </palabra>
<palabra lema="cerdo" cat="nombre" gen="masc" num="plural"> **cerditos** </palabra>
</oración>

Otra posibilidad de codificación, que puede ser bastante práctica desde diferentes puntos de vista, es utilizar algún tipo de formato tabular donde cada fila corresponde a un dato (puede ser una unidad lingüística particular, como una palabra o una oración) y cada columna a una categoría de clasificación (a una variable). Es el formato de salida estándar de analizadores automáticos como TreeTagger o FreeLing. En el ejemplo (2) se ha situado cada palabra en una fila y en las columnas se han anotado lema, clase palabra y relaciones sintácticas de dependencia. Un formato llamado CONLL-U, similar al de (2) pero con diez columnas, es el utilizado por universaldependencies.org, que contiene corpus de más de un centenar de lenguas analizadas sintácticamente con el mismo esquema de anotación.

(2)

ID	TOKEN	LEMMA	POS	HEAD	DEPREL
1	El	el	DET	2	det
2	cuento	cuento	NOUN	0	root
3	de	de	ADP	6	case
4	los	el	DET	6	det
5	tres	3	NUM	6	nummod
6	cerditos	cerdo	NOUN	2	nmod

El formato de filas y columnas también es adecuado para anotar a posteriori datos extraídos de corpus (por ejemplo, concordancias) añadiendo en columnas adicionales los rasgos que necesitemos para una investigación particular.

Combinando tablas de anotación, podemos construir una base de datos relacional, un formato que, según Davies (2005), presenta numerosas ventajas para la anotación de corpus, y que es lo que lo que utiliza el propio Davies en el trasfondo de las distintas versiones de *Corpus del Español* y otros corpus gestionados por él. Y es también, por ejemplo, el formato que se utiliza en BDS y ADESSE (García-Miguel et al. 2010).

Finalmente, el proceso de compilación, codificación y anotación de corpus debería culminar en un sistema de consulta y explotación. En los corpus puestos a disposición pública, lo importante no son tanto los formatos particulares utilizados para codificación y anotación del corpus como que los criterios de construcción del corpus estén claros y que el sistema de consulta nos permita explorarlo utilizando diferentes criterios de búsqueda. Un buen sistema de explotación de corpus nos permite refinar las búsquedas combinando criterios y, además, nos proporciona una clasificación inicial de los datos, que luego podrá refinar el usuario. Puede verse en De Benito Moreno (2019) un buen

análisis comparativo de las posibilidades que ofrece la interfaz web de trece corpus del español desde la perspectiva del usuario lingüista.

4. ¿Qué hacer con un corpus? Lo que (mejor) nos pueden ofrecer los corpus a los lingüistas

Con un corpus se pueden hacer múltiples cosas y se ha demostrado su utilidad casi en cualquier campo de los estudios del lenguaje. El desarrollo de la lingüística de corpus a partir de los años 70 fue impulsado por su utilidad para la lexicografía, la enseñanza de lenguas, la traducción y otros campos de la lingüística aplicada. Sin embargo, en este artículo y en este apartado nos centraremos en la relevancia de los datos de corpus para la lingüística teórica (en contraposición a la aplicada), esto es para la descripción y explicación de la estructura interna de las lenguas, en particular del léxico y la gramática. La lingüística de corpus, como estamos viendo, nos permite contar con datos naturales y nos proporciona herramientas para manejar una cantidad ingente de datos. Ahora bien, como toda herramienta y como toda fuente de datos, la lingüística de corpus nos inclina a fijarnos más en unas cosas que otras y a concebir el estudio del lenguaje de unas maneras mejor que de otras. Con carácter general, podemos decir que lo que mejor nos permite la lingüística de corpus es documentar si cierto fenómeno ocurre en el corpus, con qué frecuencia ocurre y cómo se distribuye en las diferentes secciones del corpus o en relación con otros elementos presentes en el corpus. Tomando la palabra (ortográfica) como unidad básica de referencia, cualquier sistema de exploración de corpus (sea un sistema de consulta online sea un sencillo programa de análisis textual como *AntConc*) nos permite obtener como mínimo alguno de los siguientes resultados:

- Listados de usos contextualizados de la palabra buscada (concordancias o ‘palabra clave en contexto’)
- Listados de las palabras del corpus (con su frecuencia)
- Listados de coapariciones: expresiones multipalabra, colocaciones, ...

Otros resultados que nos ofrecen algunos corpus, tales como palabras clave de un texto o conjunto de textos, tesauros, o resúmenes de la distribución de una palabra en el corpus, no son sino resultados derivados de los anteriores. En corpus anotados también podemos hacer lo mismo con elementos gramaticales, esto es buscar y obtener muestras de uso contextualizado y frecuencias de aparición y coaparición de cualquier categoría gramatical anotada y de patrones sintagmáticos, no solo de palabras individuales. Veremos a continuación con algo más de detalle los tipos básicos de explotación.

4.1. Documentar ocurrencias de uso

Para la persona que consulta un corpus, el procedimiento seguido en el análisis de corpus empieza normalmente por extraer de un corpus suficientemente grande fragmentos textuales que muestren fenómenos lingüísticos particulares y el objetivo deseable es que el sistema de consulta nos devuelva todos los casos y solo los casos del fenómeno investigado. En este sentido, uno de los usos más comunes de un corpus es la obtención de concordancias, que consisten en un listado de ocurrencias, presentado típicamente con una línea para cada ocurrencia, con la unidad buscada en el centro precedida y seguida

del contexto inmediato hasta cierto número de palabras. Además, encontraremos en las concordancias alguna referencia que identifica el ejemplo y, posiblemente, algún enlace que permita acceder a un contexto más amplio (a veces también a la propia grabación en el caso de corpus orales o multimedia). Por ejemplo, en (3) tenemos las primeras líneas resultantes de la búsqueda de ocurrencias de la palabra *ahorita* en Ameresco:

(3) *Ahorita* en corpus Ameresco

conversación	hablante	
HAV_046_02_17	B	pero es que dijiste ahorita que no ibas a ir
HAV_046_02_17	B	[pero] mira en esta misma premiación ahorita vamos a ponerlo
HAV_048_02_17	A	¿no fueron ahorita ? a las ocho fueron ocho ocho y pico

En los casos más simples se buscarán cadenas de caracteres que se corresponden con (partes de) palabras o frases, con lo que podemos encontrar usos contextualizados de, por ejemplo, *ahorita*, *cuento*, *hermoso*, *democracia*, *alberca*, *delante mío*, etc. En consultas un poco más complejas, se pueden buscar patrones de cadenas de caracteres con comodines o lo que se conoce como *expresiones regulares* (es decir, una especie de fórmulas en las que ciertos símbolos pueden representar clases de caracteres). Por ejemplo, en muchos corpus buscando “*hermos**” encontraremos las formas *hermoso*, *hermosa*, *hermosos*, *hermosas* y también *hermosura*, etc. Si el corpus está lematizado y etiquetado podremos buscar de la misma manera en las anotaciones correspondientes, por ejemplo, el sustantivo *cuento* (sin incluir formas del verbo *contar*), el adjetivo *hermoso* o la construcción formada por un adverbio seguido de un posesivo (tipo *delante mío/mía*). Y las búsquedas podremos hacerlas sobre todo el corpus o sobre una sección/subcorpus particular: por ejemplo, podemos buscar y comparar el uso de un taco como *joder*, en textos orales o escritos, en entrevistas frente a conversaciones, en producciones de hombres o de mujeres, en diferentes rangos de edad, etc.

Dependiendo de los niveles de anotación del corpus y del diseño del sistema de consulta, podremos refinar en mayor o menor grado las búsquedas y combinar diferentes criterios. Por ejemplo, la aplicación de consulta del CORPES, gracias al sistema de catalogación de documentos y a la lematización y etiquetación morfosintácticas, nos permite buscar ocurrencias del sustantivo *cuento* [lema] seguido de un adjetivo cualquiera [categoría del contexto] en textos orales procedentes de España [subcorpus]. Como resultado, obtenemos 41 casos en 33 documentos en la versión 0.94 de CORPES (Figura 1).

The screenshot shows the 'Corpus del Español del Siglo XXI (CORPES)' interface. At the top, there are navigation tabs: 'Concordancias', 'Comparaciones', 'Configuración', 'Ayuda', 'Estadística', 'Modo de cita', 'Sugerencias', and 'Preguntas frecuentes'. The main search area is titled 'Lema' and 'Forma', with filters for 'Clase de palabra' (set to 'sustantivo'), 'Número', 'Género', and 'Tipo'. Below this, there are options for 'Proximidad' (Distance, Interval, Left, Right, or Both) and 'Subcorpus' (Title, Author, Date, Origin, Medium, Typology, Sex, Age Group, Education Level). A list of 41 concordance results is displayed, each with a checkbox, a date, a language code, and a text snippet containing the word 'cuento' and its concordance with an adjective. At the bottom, there are options to 'Imprimir', 'Exportar', 'ExportarWord', and 'Exportar TSV', along with a page indicator '2 de 3 Ir a página: [] []'.

Figura 1. Concordancias del sustantivo *cuento* seguido de adjetivo en CORPES (subcorpus: medio oral y origen España)

El siguiente paso para el investigador consiste en observar los datos recogidos, ordenarlos y clasificarlos con diferentes criterios y elaborar hipótesis sobre las relaciones entre los datos observados. Es decir, se estudian las concordancias en busca de patrones regulares que se repiten en los ejemplos. Para ello suele ser muy útil la posibilidad de ordenar el listado por diferentes criterios, entre ellos el de ordenar por el pivote (esto es, las palabras ortográficas resultantes de la búsqueda) o por la primera palabra a la derecha o a la izquierda de la palabra buscada, de manera que aparezcan juntas algunas combinaciones que se repiten. En el ejemplo de la Figura 1 quizá fuera interesante ordenar por la primera palabra a la derecha para ver por orden alfabético qué adjetivos se registran en combinación con el sustantivo *cuento*. También podemos ordenar según alguno de los criterios de clasificación de los textos (fecha, autor, medio, etc.).

En el proceso de análisis lingüístico basado en corpus suele ser imprescindible un procesado manual de los ejemplos, es decir, una anotación de cada una de las ocurrencias utilizando criterios específicos de la investigación particular que se esté llevando a cabo. Por eso, nos interesa que el sistema de consulta permita la exportación de los resultados en un formato (.csv o .tsv) que sea fácilmente importable a una hoja de cálculo o a algún otro sistema de gestión de datos donde anotar en columnas adicionales las propiedades variables que atribuimos a cada instancia de uso.

Más arriba hemos mencionado que una propiedad destacable de la lingüística de corpus es la exhaustividad, lo que nos obliga a tener en cuenta todos los casos recogidos en el corpus según los criterios previamente establecidos y no hacer una selección de conveniencia. Puede ocurrir que estemos usando un corpus grande para investigar un

fenómeno muy frecuente, como suele ser el caso de la mayoría de los fenómenos gramaticales, y que sea imposible observar todos los ejemplos uno a uno. Pero, en general, lo que nos interesa en lingüística de corpus no suele ser el describir uno a uno todos los ejemplos encontrados sino encontrar tendencias cuantitativas generales. Para eso puede ser suficiente con la anotación que tenga el corpus. Si, por el contrario, debemos observar en los ejemplos del fenómeno analizado propiedades que no están previamente codificadas y la cantidad de ocurrencias es excesiva, lo que podemos hacer es tomar una muestra manejable que sea representativa del total. Algunos sistemas de consulta de corpus pueden proporcionar una selección aleatoria que permite sacar conclusiones sobre tendencias generales significativas.

4.2. Frecuencias

En la exploración de corpus, junto con las muestras de uso, que podemos analizar cualitativamente, obtendremos siempre datos de frecuencia. Se ha propuesto definir la lingüística de corpus como un conjunto de investigaciones que se pueden formular en términos de la distribución condicional de los fenómenos lingüísticos en un corpus (Stefanowitsch 2020: 56). De este modo, la lingüística de corpus se vincula necesariamente con los métodos cuantitativos, y con la estadística como disciplina que proporciona herramientas para el análisis cuantitativo. Como destaca Gries (2009: 11), en un corpus no hay ni significado ni funciones, solo cadenas de caracteres codificados informáticamente, por lo que lo único que realmente nos proporcionan los corpus son frecuencias de ocurrencia (cuánto aparecen ciertas palabras, construcciones gramaticales, etc. en un corpus particular) y frecuencias de coocurrencia (cuánto aparecen ciertas palabras en una construcción gramatical particular, etc.). Y le queda al investigador la tarea de interpretar esas frecuencias en términos significativos o funcionales.

Si lo único que podemos observar en el análisis de corpus son frecuencias de ocurrencias y coocurrencias en textos, lo relevante para nuestras descripciones lingüísticas es que al utilizar estos datos no nos estamos preguntando si algo es o no es posible en una lengua (menos aún, si es correcto o incorrecto), si acaso nos preguntaríamos si puede o no documentarse en el uso registrado en ese conjunto de textos. Por consiguiente, tampoco nos ponemos como objetivo la formulación de leyes que se cumplen necesariamente, en las que un fenómeno implica necesariamente otro. Lo que constatamos en el trabajo con corpus es que los elementos del lenguaje o sus combinaciones tienen una frecuencia variable y que la frecuencia relativa de cualquier unidad puede variar en relación con múltiples factores contextuales (registros, estilos, grupos sociales, dialectos geográficos, canal, diacronía, etc.) y también en relación con la coaparición de otras unidades textuales. En consecuencia, al tomar necesariamente en cuenta esa variación, la lingüística de corpus no opera con oposiciones discretas (sí o no), sino con gradaciones entre lo más o menos frecuente y, por tanto, necesita trabajar con métodos estadísticos.

Tomemos un ejemplo simple: lo que en España llamamos *ordenador* recibe en otros países hispanohablantes el nombre de *computadora*. Pero no se trata simplemente de que en la variedad *x* se utilice una forma y en la variedad *y* otra. En la Tabla 2. vemos los resultados globales que nos da el CORPES para seis grandes zonas hispanohablantes.

Zona	ORDENADOR		COMPUTADORA	
	Frec	Fnorm.	Frec	Fnorm.
España	10.636	87,86	549	4,53
México y Centroamérica	535	8,40	4.250	66,75
Río de la Plata	291	6,30	3.698	80,18
Andina	236	8,76	1.249	46,40
Antillas	236	10,48	942	41,85
Caribe continental	278	6,77	842	20,52

Tabla 2. Frecuencias absolutas y relativas (normalizadas por millón) de los lemas “ordenador” y “computadora” en el *Corpus del español del siglo XXI* (CORPES) v 0.94

Para analizar este ejemplo, debemos fijarnos en las frecuencias relativas o normalizadas (columna *Fnorm*), en este caso ocurrencias por cada millón de palabras, porque las frecuencias absolutas (columna *Frec*) lógicamente pueden depender del tamaño de cada sección del corpus. Con ello constatamos que, efectivamente, el uso de *ordenador* es mucho mayor en España que en otras zonas, mientras que el uso de *computadora* es mucho menor; pero que, en cualquier zona, podemos registrar usos de cualquiera de las dos formas. Por tanto, no hay una asociación categórica de cada zona con un lexema u otro, sino una preferencia por uno sin exclusión del otro. En este caso las diferencias son marcadas e indudablemente significativas. Pero con carácter general, y más aún en casos dudosos donde las diferencias no son tan claras, conviene acompañar observaciones de este tipo con pruebas estadísticas que muestren que las tendencias son significativas.

Lo que en el anterior ejemplo son diferencias geográficas en el léxico vale también para la variación diacrónica, social o de registro tanto en el léxico como en la gramática. En los estudios diacrónicos, por ejemplo, podemos buscar la primera o la última ocurrencia documentada de una forma lingüística, pero más interesante aún es trazar cómo aumenta o disminuye a lo largo del tiempo su frecuencia de uso. Para ello es importante, de nuevo, que comparemos frecuencias normalizadas y no frecuencias absolutas. Lamentablemente, ni CREA ni CORDE proporcionan frecuencias relativas y es difícil recuperar los datos para calcularlas uno mismo, aunque se han propuesto métodos alternativos de cálculo (Molina Salinas y Sierra Martínez 2015). En esta faceta concreta, los gráficos de CdE-hist, como el de la Figura 2, tienen la ventaja de que comparan frecuencias relativas (por millón) y no frecuencias absolutas.

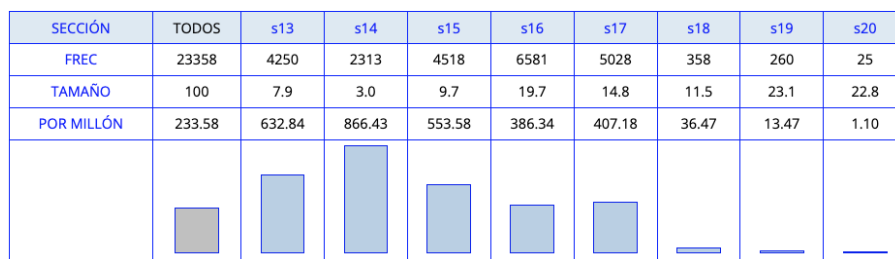


Figura 2. Frecuencias absolutas y relativas de *ahora* en CdE-hist, por siglos

Las variaciones de frecuencia aparecen por todas partes en el sistema lingüístico y dan lugar a algunos principios generales interesantes. En toda lengua hay elementos que son muy frecuentes y elementos que son muy raros, de acuerdo con la ley de Zipf, la cual establece relación inversa entre la frecuencia de una palabra y su rango en el orden de frecuencias. Esto quiere decir que, si ordenamos las palabras de cualquier corpus de más a menos frecuente, en los primeros puestos tendremos palabras de uso muy frecuente y

rápidamente van bajando las frecuencias de cada unidad hasta el punto de que encontraremos solo un ejemplar de alrededor de la mitad de las unidades. Por otro lado, sea cual sea el corpus utilizado, en los primeros puestos de la lista encontraremos siempre palabras gramaticales de muy alta frecuencia, como observamos en la Tabla 3, que contiene las palabras más frecuentes de diferentes corpus del español.

	CORPES		EsTenTen18		ARTHUS		ESLORA	
1	de	63.009	de	59.688	de	50.449	y	39.307
2	la	39.785	la	33.503	la	36.323	de	34.228
3	que	31.472	que	26.244	que	34.077	no	31.988
4	el	29.502	el	24.865	y	28.141	que	28.726
5	y	27.901	en	24.695	el	25.081	a	27.753
6	en	27.551	y	24.560	a	24.727	el	23.756
7	a	21.456	a	17.400	en	22.816	la	21.868
8	los	16.367	los	13.911	no	17.884	en	18.158
9	se	12.990	del	10.468	los	12.718	es	15.818
10	un	11.411	se	10.194	se	12.426	sí	14.863

Tabla 3. Las diez palabras más frecuentes en cuatro corpus del español.
 Frecuencias relativas (por millón)

Las diferencias que se observan en la Tabla 3. entre unos corpus y otros son muy poco significativas. De esos cuatro corpus, el que parece diferenciarse algo más de los otros es ESLORA, quizá por estar constituido por muestras de español oral mientras que los otros tres son mayoritariamente de español escrito. Lo cual nos indica una vez más que en la frecuencia de las formas en un corpus podemos encontrar algunas constantes (por ejemplo, la alta frecuencia de elementos gramaticales) junto con variabilidad debida a múltiples factores, tales como el medio oral o escrito, la procedencia geográfica, la variación diacrónica, etc.

Si observamos muestras relativamente pequeñas o textos individuales, lógicamente lo que más puede variar son los elementos léxicos, que están mucho más condicionados por la temática del texto. De los listados de frecuencia deriva un producto típico de la lingüística de corpus consistente en la obtención automática de las palabras clave de un texto o conjunto de textos, para lo que existen dos procedimientos básicos: o bien seleccionar las palabras más frecuentes de entre las de contenido léxico una vez eliminadas las palabras gramaticales, o bien seleccionar las que muestren una diferencia estadísticamente significativa respecto de la frecuencia relativa en un corpus más amplio que se toma como referencia. Para representar estos datos es práctica común el uso de “nubes de palabras”, un tipo de gráfico que destaca las palabras clave de un texto, asignando mayor tamaño a las de mayor frecuencia como en la Figura 3:



Figura 3. Nube de palabras de un borrador de este mismo artículo

La ley de Zipf también tiene otras repercusiones prácticas para el análisis de corpus: en cualquier corpus obtendremos muchos ejemplares (a veces, más de los que somos capaces de analizar en detalle) de las unidades más frecuentes; pero apenas encontraremos ejemplares de las menos frecuentes. Si, por ejemplo, estamos haciendo un diccionario basado en un corpus cerrado, es probable que de muchas palabras apenas encontremos ejemplos en el corpus. Y, debido precisamente a la ley de Zipf, duplicar el tamaño del corpus no implica ni que dupliquemos el número de palabras diferentes registradas ni que se duplique el número de ejemplares de cada palabra registrada.

4.3. Frecuencias de coaparición: de colocaciones a perfiles combinatorios

Junto con las frecuencias de las unidades y su distribución en relación con diferentes parámetros de variación, la otra gran dimensión de análisis en lingüística de corpus son las relaciones de co-ocurrencia o coaparición, que tienen que ver con cómo se combinan unas unidades con otras en los textos. Sabemos que el lenguaje es un sistema combinatorio en el que los signos básicos, palabras o morfemas se unen para formar estructuras complejas: frases, oraciones, textos. Y sabemos, también, que no todas las combinaciones imaginables son posibles y que para separar lo que es posible en una lengua de lo que no lo es se elaboraron en lingüística los conocidos conceptos de gramaticalidad y de aceptabilidad. Pues bien, como ya hemos mencionado en el apartado anterior, la lingüística de corpus no tiene ni puede tener como objetivo determinar lo que es posible en una lengua, que es algo que no se puede extraer de ningún corpus, sino estudiar qué es lo que está documentado en un corpus y con qué frecuencia. Por tanto, en el caso de las coocurrencias, no se tratará de encontrar combinaciones posibles o imposibles sino de combinaciones más o menos probables.

En la observación de coapariciones lo que se revela es la dialéctica entre lo que Sinclair (1991) ha formulado como el principio de la elección abierta [“the open choice principle”] frente al principio de idiomática [“the idiom principle”]. Por un lado, tenemos estructuras sintácticas como V + N en las que teóricamente podemos escoger cualquier nombre o cualquier verbo en las posiciones correspondientes. Por otro lado, los usuarios de la lengua tienen a su disposición un número grande de combinaciones ya construidas o semiconstruidas, que se repiten de unos textos a otros y aparecen frecuentemente en el corpus. Son fórmulas prefabricadas que no tienen por qué crearse *ex novo*, aunque su significado global pueda ser composicional, como en los sintagmas libres, o no serlo, como en las locuciones. En esta línea, la lingüística de corpus ha destacado que existen límites difusos entre las combinaciones fijas, las combinaciones preferentes más o menos restringidas y las combinaciones más abiertas con significado composicional.

Para estudiar las combinaciones de palabras, incluso en corpus que no tienen ningún tipo de anotación gramatical, se han desarrollado ciertas técnicas para resumir automáticamente la distribución de una unidad, entre las cuales destacan la búsqueda de n-gramas y el estudio de colocaciones. Los n-gramas (bigramas, trigramas, etc.) son cadenas de dos o más palabras consecutivas de uso frecuente. Es fácil generar automáticamente listas con todas las combinaciones de dos palabras, tres palabras, etc. y ordenarlas por su frecuencia. Las combinaciones que más se repiten son en buena medida fórmulas prefabricadas, aunque no necesariamente son locuciones, pues su significado global puede ser composicional o no serlo. Por ejemplo, en un corpus de recetas de cocina de elaboración propia encontramos que los trigramas más frecuentes son, por este orden,

aceite de oliva, un poco de y salsa de tomate. Esto nos proporciona información en parte extralingüística sobre qué ingredientes solemos utilizar para cocinar, pero también lingüística sobre expresiones que aprendemos y usamos como combinaciones prefabricadas en determinados contextos discursivos.

En lingüística de corpus, se llama *colocación* a la relación entre palabras que aparecen juntas con mayor frecuencia de lo esperable por pura casualidad. En este caso no se trata solo de palabras que aparecen una al lado de otra, como en los n-gramas, sino que se busca en las proximidades, entendiéndose por proximidad una “ventana” de *n* palabras a izquierda o derecha de la palabra buscada de tamaño variable y que suele fijarse en un máximo de 5 palabras a la izquierda o a la derecha. Y en ese margen lo que se buscan no son simplemente coapariciones frecuentes, sino que la coaparición sea estadísticamente significativa, comparando la frecuencia observada con la teóricamente esperable a partir de la frecuencia total que tienen en el corpus las palabras combinadas. Por ejemplo, si buscamos en CORPES sin ningún tipo de restricción cuáles son las palabras que aparecen más frecuentemente en las proximidades de un verbo como *vencer* encontraremos que las más frecuentes son el artículo *el* y las preposiciones *a* y *de*, que son las palabras más frecuentes en todo el corpus, y por tanto no puede decirse que hayan sido atraídas en particular por ese verbo. En cambio, una palabra como *timidez* tiene más probabilidades de aparecer en el entorno de *vencer* que en otras partes del corpus, por lo que podemos decir que *vencer* la atrae y que *timidez* es un colocado de este verbo.

Existen diferentes medidas estadísticas de asociación (Church y Hanks 1990) que comparan de distintas maneras la frecuencia observada en una colocación con la frecuencia esperada. El buscador de CORPES ofrece varias de estas medidas para elegir cómo ordenar los resultados. En la Tabla 4 se han extraído manualmente los nombres abstractos que coaparecen más con el verbo *vencer* y se han ordenado por MI (‘mutual information’), que utiliza una fórmula que divide frecuencia observada y esperada. Otras fórmulas de cálculo (t-score, LL-simple) situarían en primer lugar la palabra *resistencia*, otorgando más peso a las combinaciones con mayor frecuencia absoluta.

	Frec	MI	LL simple	T score
<i>timidez</i>	84	9,3	399,37	9,16
<i>repugnancia</i>	23	8,85	102,97	4,79
<i>reticencia</i>	38	8,83	169,76	6,16
<i>obstáculo</i>	226	8,5	964,74	15,03
<i>adversidad</i>	38	8,35	158,62	6,16
<i>pereza</i>	34	8,2	138,85	5,83
<i>escollo</i>	24	8,2	98,06	4,89
<i>cansancio</i>	120	8,06	480,04	10,95
<i>resistencia</i>	313	7,94	1230,43	17,63
<i>dificultad</i>	163	6,28	478,38	12,61

Tabla 4. Nombres abstractos que aparecen en las proximidades del verbo VENCER en el *Corpus del español del siglo XXI (CORPES)*. Elaboración propia

En la exploración de CORPES para buscar las coapariciones del verbo *vencer* de la Tabla 4, antes de la extracción manual de los nombres abstractos, el primer criterio (MI) situaba en los primeros puestos de la lista una serie de resultados numéricos (tipo “2-1”) y una serie de nombres de equipos de la NBA (como “Mavericks”), que aparecen muy raramente en ese corpus, pero de manera frecuente en las proximidades del verbo *vencer*.

Los otros criterios de ordenación de CORPES sitúan en las primeras posiciones palabras gramaticales muy frecuentes. La moraleja del ejemplo es que si queremos obtener resultados significativos sobre colocaciones en un corpus no quedará más remedio que evaluar los resultados que proporciona cada fórmula de cálculo. De las diferentes fórmulas alternativas, el cálculo de “información mutua” (MI) suele ofrecer buenos resultados si se descartan los casos de menor frecuencia.

La tarea de encontrar colocaciones significativas se verá facilitada enormemente si contamos con un buen sistema de exploración de corpus que nos permita ajustar los parámetros de búsqueda. Por ejemplo, el sistema de búsqueda del *Corpus del español*, además de fijar unos valores mínimos de frecuencia y de relevancia, permite entre otros ajustes seleccionar la ventana de búsqueda y la categoría del colocado, de modo que podemos buscar nombres que aparecen en hasta tres posiciones a la derecha de *vencer*, con lo que previsiblemente obtendremos objetos directos típicos de ese verbo: las primeras posiciones son ocupadas por *repugnancia*, *obstáculo*, *resistencia* y *dificultad*; lo que se corresponde bastante bien con nuestras intuiciones sobre la combinatoria de este verbo. Con todo, esto no nos libra del proceso de evaluación. En las primeras posiciones de la lista aparecían también otros sustantivos como *batalla*, *lucha* o *plazo*, que no solo pertenecen a otros campos semánticos, sino que mantienen diferente relación sintáctica con el verbo: complemento de lugar (*vencer en la batalla/lucha*) o sujeto (*vencer el plazo*). Sería deseable que los procedimientos de exploración de colocaciones fueran sensibles a la estructura sintáctica y semántica (por ejemplo, nombres abstractos que ocurren como objeto directo de un verbo dado), pero esto está supeditado a la existencia de corpus anotados sintácticamente.

Los corpus y sistemas de consulta más difundidos (incluyendo el procedimiento de elaboración de “*word sketches*” por parte de la plataforma *Sketch Engine*, más sofisticado de lo habitual pero basado en una gramática algo rudimentaria) nos ofrecen un sucedáneo basado en la categoría morfosintáctica de las palabras a la derecha y a la izquierda de la palabra nodo. Aunque presenta limitaciones, podemos obtener igualmente una muy útil visión general de la distribución de una palabra dada. La Figura 4 nos muestra parte de un *word sketch* del verbo *vencer* con las palabras que aparecen típicamente junto a él como modificadores (adverbios), objetos (nombres a la derecha del verbo o nombres modificados por el participio pasivo) u elementos coordinados.



Figura 4. *Word Sketch* del verbo VENCER en el corpus esTenTen18 (Sketch Engine)

A este respecto, cabe señalar que el concepto de colocación utilizado en lingüística de corpus, basado en la frecuencia observada en los textos, ha sido criticado con el argumento de que combinaciones frecuentes como la de cierto verbo con cierto sustantivo (por ejemplo, *leer + libro*) no nos proporcionan propiedades lingüísticas de esas palabras sino propiedades externas de las entidades del mundo a las que se aplican ciertas acciones. Esa es la tesis de Bosque (2001), quien propone usar el término colocación para referirse solo a combinaciones léxicas restringidas idiomáticamente, como por ejemplo *vino tinto* (y no *vino rojo*), pero no para referirse a cualquier combinación frecuente. Sin embargo, las colocaciones, entendidas como en lingüística de corpus como coapariciones más frecuentes de lo esperable por el azar, son imprescindibles para conocer el significado y uso de las palabras. Aportan conocimiento idiomático, normas de uso específicas de una lengua particular, como, por ejemplo, que el *vino es blanco o tinto*, no amarillo y rojo. Así mismo, aportan conocimiento social-discursivo: qué cosas se suelen decir de qué; por ejemplo, qué adjetivos y nombres valorativos se usan en la prensa alrededor de la palabra ‘musulmán’ y, por tanto, qué prejuicios y actitudes se transmiten hacia los musulmanes en el discurso publicado (Baker et al. 2013). Y, ciertamente, aportan también conocimiento de la realidad designada por las palabras, como que *leer* se combina frecuentemente con *libro*; algo que por supuesto puede cambiar si cambia la realidad: así, en CdE-hist los nombres que se colocan más frecuentemente a la derecha del verbo *leer* son *libro, carta, periódico y novela*; pero en el más reciente CdE-web la misma búsqueda da como resultado *libro, artículo, comentario y blog*. En definitiva, por mucho que estén determinadas por la realidad designada, son esas coapariciones las que utilizamos para interpretar y aprender el significado de las palabras.

Como en el corpus no hay significado como tal, solo hay cadenas de caracteres y frecuencias de aparición y coaparición, son esas frecuencias las que nos están informando sobre cuál es el significado de las palabras y en qué se parecen y en qué se diferencian. La lingüística de corpus se adhiere así a la hipótesis distribucional (elaborada por Bloomfield y seguidores, en particular por Z. S. Harris) según la cual las entidades lingüísticas que tienen una distribución similar tienen significados similares y, viceversa, las diferencias de significado se reflejan en diferencias de distribución. De esta manera, podemos hacer con el análisis de corpus una semántica basada en el uso utilizando métodos estadísticos. En el procesamiento del lenguaje (PLN) o lingüística computacional, este principio permite desarrollar métodos automáticos a partir de corpus (anotados o no) para agrupar palabras semánticamente similares porque tienen distribuciones similares, infiriendo de ello semejanzas y diferencias de significado (Martí Antonín 2018).

De ese principio general, que las diferencias y semejanzas de significado se manifiestan en diferencias y semejanzas de distribución, deriva también el concepto de *perfil combinatorio*, que da lugar a un método derivado usado en semántica lingüística basada en corpus consistente en seleccionar muestras de la palabra o palabras analizadas y anotar manualmente rasgos formales y funcionales que presentan los ejemplos contextualizados de la muestra y luego aplicar métodos estadísticos multifactoriales de agrupación y discriminación que permitan describir la estructura de un campo semántico, o la de una palabra polisémica o mostrar las diferencias entre cuasi-sinónimos a partir de su perfil combinatorio [*Behavioral Profile*] (Gries y Divjak 2006, Gries 2010a, Glynn y Robinson 2014). La Figura 5 muestra en una estructura arbórea las semejanzas relativas entre verbos en ruso del campo semántico “intentar” de acuerdo con este método.

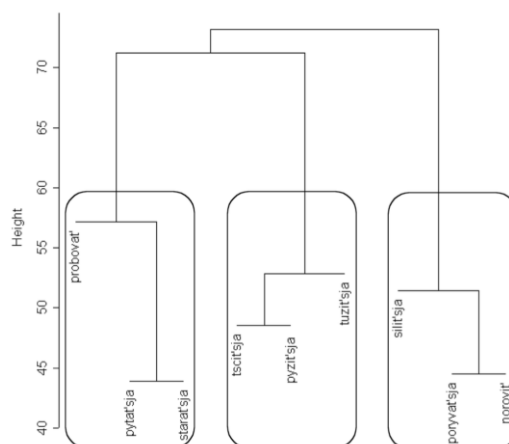


Figura 5. Dendrograma de verbos intencionales en ruso según su perfil combinatorio (Divjak y Gries 2006: 38)

4.4. Patrones sintagmáticos

En lo que concierne a la relación de la lingüística de corpus con los estudios sintácticos, como se ha comentado un poco más arriba, muchos corpus contienen anotación morfosintáctica de las palabras, pero es raro que los grandes corpus contengan una anotación sintáctica detallada, algo reservado a corpus relativamente pequeños con anotación total o parcialmente manual. La anotación sintáctica de corpus, como por ejemplo la integrada en ADESSE, permite investigar en detalle fenómenos sintácticos como la marcación variable del objeto (García-Miguel 2015), esto es, la presencia variable de la preposición *a*, la duplicación pronominal o la elección de caso acusativo (*lo*) o dativo (*le*) con objetos directos e indirectos. Fenómenos de este tipo pueden investigarse también en un corpus mínimamente anotado, pues con búsquedas relativamente simples podemos documentar el uso y distribución de cadenas de caracteres como *le* y *lo*.

En corpus con anotación morfosintáctica podemos explorar, como se ha indicado antes, no solo palabras y frases particulares sino también categorías gramaticales en sí mismas –por ejemplo, cuál es la frecuencia relativa de nombres y verbos en un (sub)corpus– o en combinación con otros elementos –por ejemplo, qué nombres aparecen a la derecha del verbo *vencer* o cuál es la distribución de la construcción Adverbio + Posesivo del tipo *detrás mío* (Eddington 2017). Ciertamente, la metodología basada en corpus no suele centrarse en reglas combinatorias muy generales, sino en la exploración de construcciones particulares de un nivel de abstracción relativamente bajo.

Algunas aproximaciones teóricas a la gramática se basan en patrones sintagmáticos que emergen del corpus. Más que en reglas o principios combinatorios abstractos de aplicación general, tales patrones combinan elementos abstractos o esquemáticos con elementos léxicos específicos o clases de elementos léxicos de manera que se difumina la distinción entre léxico y gramática. En esta línea podemos mencionar la *Pattern Grammar* de Hunston y Francis (1998, 2000), el '*Corpus Pattern Analysis*' de P. Hanks (Hanks y Pustejovski 2005) y otros trabajos que destacan que la descripción de la combinatoria de una unidad léxica debe incluir junto con la valencia sintáctico-semántica también información colocacional (Butler 2001). La técnica de análisis consiste en observar instancias de uso en corpus y agrupar las líneas de concordancia según patrones

sintagmáticos semánticamente motivados, de modo que los significados se asocian con contextos prototípicos. En el caso de los verbos, los patrones no incluyen simplemente la estructura valencial sino también detalles contextuales potencialmente relevantes para el significado, como puede ser la presencia de ciertos nombres o clases de nombres, de cierta preposición o incluso la presencia o ausencia de determinante junto al nombre complemento (por ej. *tener lugar* es distinto de *tener un lugar*). En (4) tenemos un ejemplo en inglés de Hunston y Francis con los patrones que son parte de la conducta típica del sentido relevante de cada verbo. En (5) tenemos parte de la entrada del verbo *preocupar* en Verbario, siguiendo el modelo de *Corpus Pattern Analysis* utilizado en el PDEV por P. Hanks.

- (4) *we tended to think of ourselves as not having to worry about those freaks down there* [‘tendíamos a pensar en nosotros mismos como si no tuviéramos que preocuparnos por esos monstruos de abajo’]

Verb patterns:

tended V to-inf

think V of n as n/-ing

worry V about n

(Hunston & Francis 1998: 64-65)

- (5) Patrones de PREOCUPAR en Verbario

1a [[Humano | Eventualidad]] *preocupar* [[a Humano]]

1b [[Humano 1]] *preocuparse* (por/de Humano 2 | Eventualidad)

El análisis llamado colostruccion, propuesto por Stefanowitsch y Gries (2003), combina el concepto de colocación con el concepto de construcción. El objetivo es observar en una construcción dada qué lexemas particulares entran preferentemente en los huecos abiertos por la construcción, fuertemente atraídos por ella. El análisis puede aplicarse tanto a construcciones muy específicas, como [N *waiting to happen*] (lit. ‘N esperando a suceder’), para las que podemos explorar en un corpus del inglés qué nombres pueden rellenar el hueco que queda libre, como a esquemas puramente abstractos como la construcción ditransitiva ([S_{agent} V O_{rec} O_{th}] en inglés, [Suj – V – OD – OI] en español), para la que podemos explorar en corpus cuáles son los verbos atraídos por ese esquema. En este último caso se comprueba que no cabe cualquier nombre ni cualquier verbo en ese esquema, sino que la construcción está claramente asociada con el verbo *dar* y otros verbos semánticamente relacionados con la transferencia, tanto en inglés como en español (Tabla 5). Stefanowitsch y Gries utilizan una prueba estadística, el test exacto de Fisher, para calcular la fuerza de la asociación [*collostructional strength*], aunque en este caso concreto la observación de frecuencias absolutas puede proporcionarnos resultados similares.

Collexeme	Coll. strength	VERBO	Frec
<i>give</i> (461)	0	<i>dar</i>	1184
<i>tell</i> (128)	1.6E-127	<i>decir</i>	598
<i>send</i> (64)	7.26E-68	<i>hacer</i>	486
<i>offer</i> (43)	3.31E-49	<i>contar</i>	308
<i>show</i> (49)	2.23E-33	<i>pedir</i>	272
<i>cost</i> (20)	1.12E-22	<i>preguntar</i>	220
<i>teach</i> (15)	4.32E-16	<i>poner</i>	144
<i>award</i> (7)	1.36E-11	<i>permitir</i>	124

Tabla 5. Verbos asociados con la construcción ditransitiva en inglés (Stefanowitsch y Gries 2003: 229) y en español (ADESSE. Elaboración propia)

La moraleja es que los esquemas sintácticos abstractos se asocian con elementos léxicos particulares. Y viceversa, los elementos léxicos no tienen solo una manera de construirse. Los verbos no tienen una valencia fija, sino que admiten diferentes esquemas construccionales y tienen una probabilidad mayor o menor de aparecer en un esquema o en otro (García-Miguel 2005), como podemos observar en el ejemplo de la Figura 6 con las frecuencias de los esquemas valenciales del verbo *vencer* en ADESSE.

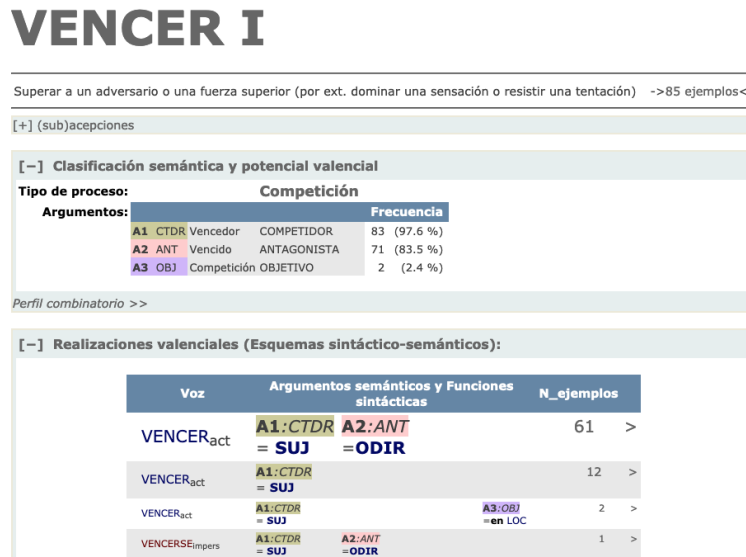


Figura 6. Esquemas sintáctico-semánticos de *VENCER* (ADESSE)

En definitiva, lo que más y mejor solemos hacer con un corpus es buscar patrones de coocurrencia, de elementos léxicos con elementos léxicos (colocaciones), de elementos y esquemas gramaticales con elementos léxicos (por ej. colostrucciones), de elementos léxicos con esquemas gramaticales (valencia), etc. Podemos generalizar el concepto de *perfil combinatorio* para referirnos al conjunto de probabilidades de coocurrencia de una unidad lingüística cualquiera con elementos léxicos y gramaticales de cualquier nivel de generalidad. Estas probabilidades varían según el contexto, por lo que deberíamos añadir factores sociales y discursivos y aplicar técnicas estadísticas de análisis multifactorial, algo que no desarrollaremos por los límites de espacio de este artículo.

4.5. Final de sección

En resumen, virtualmente cualquier aspecto del léxico y de la gramática de las lenguas puede estudiarse en un corpus. Pero en la práctica es lógico que existan ciertos sesgos motivados por la relativa facilidad para digitalizar textos, anotarlos y explorarlos con ayuda de la informática. Una tendencia general es la de focalizar el interés en la frecuencia de unidades y patrones que son directamente observables en el corpus con ayuda de ordenador. Pero tales observaciones necesitan acompañarse por parte del lingüista de procesos de abstracción tanto en la clasificación como en la interpretación de los resultados obtenidos. Según Barlow (2011: 7-8) las principales aportaciones de la lingüística de corpus a la investigación lingüística han sido tres: (i) destacar la

importancia de las co-apariciones, (ii) focalizarse en la frecuencia y, por tanto, en formas típicas de expresión más que en formas posibles, y (iii) cuantificar el rango de variación en las lenguas. Por otro lado, los estudios guiados por corpus suelen presentar ciertos rasgos comunes (adaptado de Barlow 2011: 11):

- prominencia de las unidades léxicas en las descripciones;
- se recurre a categorías más abstractas o esquemáticas (categorías, relaciones) solo en caso necesario;
- las semejanzas y diferencias de significado se asocian con la distribución;
- los procesos de interpretación y la conexión con aspectos cognitivos, sociales, pragmáticos o discursivos no siempre se establecen de manera sistemática.

5. Lingüística de corpus y teoría(s) lingüística(s)

La lingüística de corpus es ante todo una metodología cuyos conceptos y resultados son compatibles con diferentes modelos lingüísticos, de modo que cualquier modelo teórico puede utilizar, en mayor o menor medida, datos de corpus. Sin embargo, es cierto que algunas perspectivas se han mostrado más proclives que otras a utilizar datos de corpus y que los datos de corpus pueden utilizarse de diferentes maneras con distinta repercusión en la teoría. Puede distinguirse (Tummers et al. 2005: 234-236) entre lingüística *ilustrada con corpus* y lingüística *basada en corpus*. La primera toma ejemplos de corpus para complementar datos y realiza análisis basados más bien en la intuición. En la segunda, la evidencia empírica y las tendencias encontradas en el uso constituyen el núcleo del análisis y definen el modelo resultante, y en estas aproximaciones basadas en corpus las técnicas estadísticas constituyen una parte esencial del análisis.

Muchos lingüistas que hacen sintaxis más o menos formal piensan que con la intuición es suficiente para elaborar modelos teóricos de lo que es posible en una lengua, o en general en las lenguas. Entre ellos se encuentran lingüistas tan prominentes como Chomsky o Tesnière. Este último afirma explícitamente que la introspección debe convertirse en una de las piezas maestras de la investigación sintáctica (Tesnière 1959: 37). Pero también hay quienes hacen sintaxis más o menos formal y han incorporado nociones de gradación, frecuencia y probabilidad basadas en la observación del uso real, por tanto, en corpus (por ejemplo, Manning 2003, Aarts 2007, Bresnan y Hay 2008).

Con todo, con lo que parece mostrar mayor compatibilidad la lingüística de corpus es con los modelos teóricos que se dicen “basados en el uso” (Barlow y Kemmer 2000), esto es, los modelos cognitivos y funcionales. Si se pretende basar el modelo en el uso hay que observar el uso real y el método que surge como primer candidato es, obviamente, la lingüística de corpus. Destacaremos en particular la conexión de la lingüística de corpus con el contextualismo británico (incluyendo la Lingüística Funcional de Halliday) y con la Lingüística Cognitiva.

Donde más se desarrolló inicialmente la lingüística de corpus fue en Gran Bretaña, en buena parte como continuación de la perspectiva contextualista de Firth (1890-1960) a quien debemos el aforismo “a una palabra se la conoce por sus acompañantes” [“You shall know a word by the company it keeps”] (Firth 1957: 179), que está en la base de su concepto de colocación. Firth influye en la llamada lingüística de corpus neo-firthiana

(Sinclair, Hunston, Teubert, ...), con su aproximación inductiva “guiada por corpus” [*corpus-driven*]; pero también influye en otros lingüistas británicos con una orientación más “basada en corpus” [*corpus-based*] (Aarts, Biber, Leech, McEnery). Esta distinción entre dos aproximaciones aparentemente contrapuestas fue formulada por Tognini-Bonelli (2001), según la cual, en las aproximaciones “basadas en corpus”, la teoría y las categorías analíticas se establecen previamente y se comprueban o ejemplifican en el corpus; mientras que, en las aproximaciones “guiadas por corpus”, la teoría y las categorías analíticas derivan de los datos, emanan del corpus, siguiendo un método puramente inductivo. Esta última perspectiva es la defendida por la propia Tognini-Bonelli y por figuras destacadas como John Sinclair (Sinclair 2004). Por el contrario, otros especialistas (McEnery et al. 2006: 8-10) aducen que la distinción entre esas dos aproximaciones se ha exagerado y que es más bien borrosa; lo que nos da pie a propugnar una metodología que recorra continuamente el camino de ida y vuelta entre teoría y datos.

La insistencia de algunos lingüistas británicos, especialmente de Sinclair, en seguir un método estrictamente inductivo que evite cualquier categoría que no emane del propio corpus les da a estas aproximaciones una apariencia atórica o carente de un modelo teórico elaborado. Por el contrario, no tiene nada de atórica la Lingüística Sistémico-Funcional de Michael Halliday, también notablemente influida por Firth. La teoría de Halliday destaca por su capacidad para integrar la explicación del sistema lingüístico en relación con el texto y el contexto. Aunque esta teoría no está necesariamente ligada al uso de corpus (pero sí se usan textos reales como ilustración de la teoría, más que oraciones inventadas), Halliday siempre defendió la relevancia teórica de la lingüística de corpus y la importancia de la frecuencia en la conformación del sistema lingüístico. Para él, el significado deriva de una red de sistemas de elecciones (de ahí el término “sistémica”) y una parte importante del significado de un término cualquiera es la probabilidad de escogerlo frente a otro(s). Así, por ejemplo, el significado de ‘negativo’ no es simplemente ‘no positivo’ sino ‘no positivo con una razón de probabilidades de uno a nueve’. Como las probabilidades varían según el contexto, las elecciones gramaticales pueden significar cosas diferentes en registros diferentes (Halliday 1991: 32-33).

La influencia de Firth se refleja también en el interés creciente por los estudios basados en corpus de las relaciones entre lenguaje, contexto, sociedad y cultura (Gabrielatos 2021).

Un campo donde la lingüística de corpus ha encontrado terreno muy abonado es el de la Lingüística Cognitiva (Tummers et al. 2005; Gries 2010b: 334-336). Frente a la dicotomía tradicional entre sistema (lengua, competencia) y uso (habla, actuación), la Lingüística Cognitiva propone que el sistema está conformado desde el principio por el uso, y que las estructuras lingüísticas emergen del uso individual y colectivo del lenguaje. Y la lingüística basada en el uso necesita cuantificación y análisis estadístico (Tummers et al. 2005: 234). En este sentido, un concepto clave para el modelo de gramaticalidad basado en el uso es el de *entrenchment* [‘consolidación mental’]: las estructuras lingüísticas nuevas se consolidan progresivamente gracias al uso repetido y las unidades estarán más o menos consolidadas dependiendo de su frecuencia de ocurrencia (Langacker 1987: 59). Este concepto de consolidación mental reformula el concepto de gramaticalidad (que además deja de ser una cuestión de sí o no, para convertirse en algo gradual) y le proporciona una base empírica que puede contrastarse con las frecuencias observadas en un corpus. Sin embargo, hay que cuidarse de establecer una relación directa entre frecuencia simple y consolidación mental (Arppe et al. 2010: 8-10). Otras facetas

en las que encaja bien la lingüística de corpus con la Lingüística Cognitiva están en el estudio de la categorización basada en ejemplares y en los efectos de prototipicidad en la polisemia (Mukherjee 2004; Glynn y Robinson 2014), nuevamente con la advertencia de que no se puede equiparar prototipicidad y frecuencia simple. Y la concepción del lenguaje como un sistema complejo que no se puede entender haciendo abstracción de factores contextuales y sociales (algo también esencial en la perspectiva cognitivista) se plasma en estudios cuantitativos basados en corpus que aplican modelos multifactoriales que incluyen explícitamente la variación social y cultural (Heylen et al. 2008).

Tanto la Gramática Cognitiva como su pariente próximo la Gramática de Construcciones (Goldberg 2003) entienden la gramática como una red de unidades convencionales que asocian formas y significados a diferentes niveles de esquematicidad y fijación. En una línea similar, perfectamente compatible con la de la Lingüística Cognitiva, tenemos trabajos como los de Bybee, Hopper y otros (Bybee y Hopper 2001; Bybee 2007, etc.), con la idea de que la gramática emerge del uso, es desde su raíz variable y probabilística y se constituye a partir de instancias específicas de uso mediante procesos de rutinización y esquematización basados en la categorización de ejemplares. En esta aproximación se destaca especialmente el papel de la frecuencia en los mecanismos de cambio y gramaticalización (Bybee 2003).

6. (A modo de) conclusión

Cualquier lingüista, casi con independencia de su campo de investigación, puede explorar corpus de distintas fuentes con la seguridad de que, en poco tiempo, podrá acumular una gran cantidad de datos relevantes (muchas veces mezclados con datos irrelevantes) y que entre ellos habrá muchos datos sorprendentes. Pero no se debe olvidar que el objetivo de cualquier investigación es avanzar en el conocimiento del objeto de estudio, no simplemente acumular datos. Y en ese objetivo deben evaluarse siempre las limitaciones que pueda tener tanto el corpus utilizado como la propia metodología basada en corpus, por lo que idealmente debería complementarse con otros métodos como los que se exponen en otros capítulos de este volumen. Como indicamos más arriba, se ha comparado la importancia para la lingüística de la informatización de datos textuales con lo que supuso el telescopio para la astronomía; pero cada herramienta tiene su función y su utilidad dependerá de lo que queramos hacer: tan ridículo como criticar un telescopio por no servir de microscopio sería el criticar la lingüística de corpus por no hacer aquello para lo que no está concebida (McEnery et al. 2006: 121).

Para conseguir adecuación descriptiva necesitamos corpus textuales representativos de la lengua o variedad de lengua objeto de estudio. También necesitamos anotaciones detalladas, pues no siempre es suficiente con los resultados que nos proporciona la búsqueda de secuencias de caracteres de palabras o de categorías morfosintácticas. Las categorías usadas en la anotación de corpus deben estar fundamentadas teóricamente y deben revisarse a la luz de la evidencia proporcionada por el propio corpus. Debemos movernos continuamente de las categorías analíticas al corpus y del corpus a las categorías analíticas. En mi opinión, se observa en las últimas décadas cierta aproximación entre los estudios más teóricos y los más empiristas, aunque probablemente aún queden muchos pasos por dar hasta llegar a la integración total. Los estudios de corpus incluyen propiedades sintácticas y semánticas cada vez más refinadas y los

estudios más teóricos están prestando progresivamente más atención a los datos empíricos y a las variaciones que se observan en el uso. Entre las principales aportaciones de la lingüística de corpus a la lingüística contemporánea está, además de la puesta a disposición de una ingente cantidad de datos, el desarrollo de métodos cuantitativos que ponen en relación la frecuencia de uso con múltiples factores co-textuales y contextuales.

7. Referencias de corpus y otros recursos lingüísticos mencionados

- ADESSE: Base de datos de Verbos, Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español. <http://adesse.uvigo.es>
- Ameresco: Corpus América y España Español Coloquial. <http://esvaratenuacion.es/>
- AnCora: Multilingual and multilevel ANnotated CORpora. <http://clic.ub.edu/corpus/>
- AntConc: [software] <https://www.laurenceanthony.net/software/antconc/>
- Auslan Corpus: Australian Sign Language Corpus.
<https://www.auslan.org.au/about/corpus/>
- BDS: Base de Datos Sintácticos del Español Actual. <http://www.bds.usc.es/>
- BNC: British National Corpus. <http://www.natcorp.ox.ac.uk/>
- Brown corpus: The Standard Corpus of Present-Day Edited American English.
<https://varieng.helsinki.fi/CoRD/corpora/BROWN/>
- BSL Corpus: British Sign Language Corpus Project. <https://bslcorpusproject.org/>
- CAES: Corpus de Aprendices de Español como lengua extranjera.
<http://galvan.usc.es/caes/>
- CdE-hist: Corpus del Español: Género/histórico.
<https://www.corpusdelespanol.org/hist-gen/>
- CdE-NOW: Corpus del Español: NOW (News on the web).
<https://www.corpusdelespanol.org/now/>
- CdE-web: Corpus del Español: web/dialectos.
<https://www.corpusdelespanol.org/web-dial/>
- CHILDES: Child Language Data Exchange System. <http://childes.talkbank.org/>
- CLUVI Parallel Corpus. <http://sli.uvigo.gal/CLUVI/>
- COLA: Corpus Oral del Lenguaje Adolescente.
<https://blogg.hiof.no/colam-esp/el-corpus-cola/>
- CORDE: Corpus Diacrónico del Español. <http://corpus.rae.es/cordenet.html>
- CORELE: Corpus Oral de Español como Lengua Extranjera.
<http://www.llf.uam.es/ESP/CORELE.html>
- CORLEC: Corpus Oral de Referencia de la Lengua Española Contemporánea.
<http://www.llf.uam.es/ESP/Corlec.html>
- CorLSE: Corpus de la Lengua de Signos Española. <https://corpuslse.es/>
- CORPES: Corpus del Español del siglo XXI. <http://web.frl.es/CORPES/>
- Corpus del Español. <http://corpusdelespanol.org/>
- Corpus Tècnic de l'IULA. <https://www.upf.edu/web/iula/corpus-eines>
- Corpus Técnico do Galego. <http://sli.uvigo.es/CTG/>
- CREA: Corpus de Referencia del Español Actual. <http://corpus.rae.es/creanet.html>
- DGS-Korpus: Deutsche Gebärdensprache Korpus.
<https://www.sign-lang.uni-hamburg.de/dgs-korpus/>
- ELAN: [software] <https://archive.mpi.nl/tla/elan>

enTenTen20: Corpus of the English web.

<https://www.sketchengine.eu/ententen-english-corpus/>

ESLORA: Corpus para el estudio del español oral de Galicia. <http://eslora.usc.es/>

esTenTen18: Spanish web corpus 2018.

<https://www.sketchengine.eu/estenten-spanish-corpus/>

FreeLing: [software] <http://nlp.lsi.upc.edu/freeling/>

IULA Spanish Treebank. http://www.iula.upf.edu/recurs01_tbk_uk.htm

NGT Corpus: Corpus Nederlandse Gebarentaal. <https://www.corpusngt.nl/>

PDEV: Pattern Dictionary of English Verbs. <https://pdev.sketchengine.eu/>

PRESEEA: Corpus del Proyecto para el estudio sociolingüístico del español de España y de América. <https://preseea.linguas.net/Corpus.aspx>

SenSem corpus: Sentence Semantics, Base de datos de semántica oracional.

<http://grial.edu.es/sensem/corpus>

Sketch Engine. <http://www.sketchengine.eu>

TEI: Text Encoding Initiative. <http://www.tei-c.org/>

TenTen corpus family. <https://www.sketchengine.eu/documentation/tenten-corpora/>

TreeTagger: [software] <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

Val.Es.Co: Corpus Valencia, Español Coloquial

<https://www.uv.es/corpusvalesco/consulta.html>

Verbario: Semántica de los verbos en contexto. <http://www.verbario.com>

8. Referencias

- Aarts, Bas. 2007. *Syntactic gradience: The nature of grammatical indeterminacy*. Oxford: Oxford University Press.
- Arppe, Antti; Gilquin, Gaëtanelle; Glynn, Dylan; Hilpert, Martin; Zeschel, Arne. 2010. Cognitive Corpus Linguistics: five points of debate on current theory and methodology. *Corpora* 5.1: 1–27.
- Baker, Paul; Gabrielatos, Costas; McEnery, Tony. 2013. Sketching Muslims: A corpus driven analysis of representations around the word “Muslim” in the British press 1998–2009. *Applied Linguistics* 34.3: 255–278.
- Barlow, Michael. 2011. Corpus linguistics and theoretical linguistics. *International Journal of Corpus Linguistics* 16: 3–44.
- Barlow, Michael; Kemmer, Suzanne, eds. 2000. *Usage-based models of language*. Stanford: CSLI.
- Biber, Douglas. 1993. Representativeness in corpus design. *Literary and Linguistic Computing* 8.4: 243–257.
- Bresnan, Joan; Hay, Jennifer. 2008. Gradient grammar: An effect of animacy on the syntax of give in New Zealand and American English. *Lingua* 118.2: 245–259.
- Bosque, Ignacio. 2001. Sobre el concepto de “colocación” y sus límites. *Lingüística Española Actual* 23.1: 9–40.
- Butler, Christopher. S. 2001. A matter of GIVE and TAKE: corpus linguistics and the predicate frame. *Revista Canaria de Estudios Ingleses* 42: 55–78.
- Bybee, Joan. 2003. Mechanism of change in grammaticization: the role of frequency. En B. Joseph y R. Janda, eds. *The Handbook of Historical Linguistics*. Oxford: Blackwell, pp. 602–623.

- Bybee, Joan. 2007. *Frequency of use and the organization of language*. New York: Oxford University Press.
- Bybee, Joan; Hopper, Paul, eds. 2001. *Frequency and the emergence of Linguistic structure*. Amsterdam: John Benjamins.
- Church, Kenneth W.; Hanks, Patrick. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics* 16.1: 22–29.
- Davies, Mark. 2005. The advantage of using relational databases for large corpora: Speed, advanced queries, and unlimited annotation. *International Journal of Corpus Linguistics* 10.3: 307–334.
- De Benito Moreno, Carlota. 2019. Los corpus del español desde la perspectiva del usuario lingüista. *Scriptum digital* 8: 1–21.
- Divjak, Dagmar; Gries, Stefan. 2006. Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2.1: 23–60.
- Dyson, Freeman J. 1997. *Imagined worlds*. Harvard University Press.
- Eddington, David. 2017. Nominalized adverbs in Spanish: The intriguing case of *detrás mío* and its cohorts. *Research in Corpus Linguistics* 5: 47–55.
- Fillmore, Charles J. 1992. “Corpus linguistics” or “Computer-aided armchair linguistics”. En J. Svartvik, ed. *Directions in Corpus Linguistics*. Berlin: Mouton de Gruyter, pp. 35–60.
- Firth, John. 1957. *Papers in Linguistics*. Oxford University Press.
- Gabrielatos, Costas. 2021. Bibliography of discourse-oriented corpus studies. <http://ehu.ac.uk/docsbiblio>.
- García-Miguel, José M. 2005. Aproximación empírica a la interacción de verbos y esquemas construccionales, ejemplificada con los verbos de percepción. *Estudios de Lingüística* 19: 169–191.
- García-Miguel, José M. 2015. Variable coding and object alignment in Spanish: A corpus-based approach. *Folia Linguistica* 49.1: 205–256.
- García-Miguel, José M.; Vaamonde, Gael; González Domínguez, Fita. 2010. ADESSE, a Database with Syntactic and Semantic Annotation of a Corpus of Spanish. En *LREC2010 - Proceedings of the Seventh International Conference on Language Resources and Evaluation*. Valletta (Malta): ELRA, pp. 1903–1910.
- Gilquin, Gaëtanelle; Gries, Stefan. 2009. Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5.1: 1–26.
- Glynn, Dylan; Robinson, Justyna A., eds. 2014. *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*. Amsterdam: John Benjamins.
- Goldberg, Adele. 2003. Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences* 7.5: 219–224.
- Gries, Stefan. 2009. *Quantitative Corpus Linguistics with R: A practical introduction*. Londres: Routledge.
- Gries, Stefan. 2010a. Behavioral profiles: a fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon* 5.3: 323–346.
- Gries, Stefan. 2010b. Corpus linguistics and theoretical linguistics: A love–hate relationship? Not necessarily... *International Journal of Corpus Linguistics* 15.3: 327–343.
- Halliday, M.A.K. 1991. Corpus studies and probabilistic grammar. En K. Aijmer y B. Altenberg, eds. *English corpus linguistics: Studies in honour of Jan Svartvik*. London: Longman, pp. 30–43.

- Hanks, Peter; Pustejovsky, James. 2005. A pattern dictionary for natural language processing. *Revue Française de Linguistique Appliquée*, 10.2: 63–82.
- Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45.1: 31–80.
- Heylen, Kris; Tummers, José; Geeraerts, Dirk. 2008. Methodological issues in corpus-based Cognitive Linguistics. En G. Kristiansen, ed. *Cognitive Sociolinguistics: Language variation, cultural models, social systems*. Berlin: Mouton de Gruyter, pp. 91–128.
- Hunston, Susan; Francis, Gill. 1998. Verbs observed: A corpus-driven pedagogic Grammar. *Applied Linguistics* 19.1: 45–72.
- Hunston, Susan; Francis, Gill. 2000. *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.
- Ide, Nancy; Pustejovsky, James, eds. 2017. *Handbook of Linguistic annotation*. Dordrecht: Springer.
- Johnston, Trevor. 2010. From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics* 15.1: 106–131.
- Kipp, Michael; Martin, Jean-Claude; Paggio, Patrizia; Heylen, Dirk, eds. 2009. *Multimodal corpora: From models of natural interaction to systems and applications*. Berlin: Springer.
- Knight, Dawn; Adolphs, Svenja. 2020. Multimodal corpora. En M. Paquot y S. Gries, eds. *A practical handbook of Corpus Linguistics*. Springer, pp. 353–370.
- Labov, William. 1972. The study of language in its social context. En J.A. Fishman, ed. *Advances in the Sociology of language*, v. 1. The Hague: Mouton, pp. 152–216.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*, Volume 1: *Theoretical Prerequisites*. Stanford: Stanford University Press.
- Leech, Geoffrey. 1992. Corpora and theories of linguistic performance. En J. Svartvik, ed. *Directions in Corpus Linguistics*. Berlin: Mouton de Gruyter, pp. 105–122.
- Manning, Christopher. 2003. Probabilistic syntax. En R. Bod, J. Hay y S. Jannedy, eds. *Probabilistic Linguistics*. Cambridge: MIT Press, pp. 289–341.
- Martí Antonín, Antonia M. 2018. Modelos de semántica distribucional. En M. Diaz-Ferro et al., eds. *Actas do XIII Congreso Internacional de Lingüística Xeral*. Universidade de Vigo, pp. 16–22.
- McEnery, Tony; Xiao, Richard; Tono, Yukio. 2006. *Corpus-based language studies: An advanced resource book*. London: Routledge.
- Molina Salinas, Claudio; Sierra Martínez, Gerardo. 2015. Hacia una normalización de la frecuencia de los corpus CREA y CORDE. *Revista signos* 48.89: 307–331.
- Mukherjee, Joybrato. 2004. Corpus data in a usage-based cognitive grammar. En K. Aijmer y B. Altenberg, eds. *Advances in Corpus Linguistics*. Amsterdam: Brill Rodopi, pp. 83–100.
- Pérez, Ania; García-Miguel, José M.; Cabeza, Carmen. 2019. Anotación de corpus para o estudo da expresión gramatical de eventos: notas sobre o deseño do proxecto RADIS. *Sensos-e* 6.1: 40–61.
- Rojo, Guillermo. 2016. Los corpus textuales del español. En J. Gutiérrez-Rexach, ed. *Enciclopedia de Lingüística Hispánica*. Oxford: Routledge, pp. 285–296.
- Rojo, Guillermo. 2021. *Introducción a la lingüística de corpus en español*. London: Routledge.

- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, John. 2004. *Trust the text: Language, corpus and discourse*. Londres: Routledge.
- Stefanowitsch, Anatol. 2020. *Corpus linguistics: A guide to the methodology*. Berlin: Language Science Press.
- Stefanowitsch, Anatol; Gries, Stefan. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8.2: 209–243.
- Stubbs, Michael. 1996. *Text and corpus analysis*. Oxford: Blackwell.
- Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. Paris: Klincksieck.
- Tognini-Bonelli, Elena. 2001. *Corpus linguistics at work*. Amsterdam: Benjamins.
- Torruella, Joan. 2017. *Lingüística de corpus: génesis y bases metodológicas de los corpus (históricos) para la investigación en lingüística*. Frankfurt am Main: Peter Lang.
- Torruella, Joan; Llisterri, Joaquim. 1999. Diseño de corpus textuales y orales. En J.M. Bleca, G. Clavería, C. Sánchez y J. Torruella, eds. *Filología e informática: Nuevas tecnologías en los estudios filológicos*. Barcelona: UAB / Ed. Milenio, pp. 45–81.
- Tummers, Jose; Heylen, Kris; Geeraerts, Dirk. 2005. Usage-based approaches in Cognitive Linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory* 1.2: 225–261.
- Valenzuela, Javier. 2022. El big data en los estudios del lenguaje. *Estudios de Lingüística del Español* 45: 241–260.

Métodos de la dialectología cuantitativa

Gotzon Aurrekoetxea
UPV-EHU
gotzonaurre@gmail.com

Resumen

La introducción de la cuantificación de la variación geolingüística ha traído consigo un espectacular auge de las publicaciones sobre la materia, que indican una renovada vitalidad de la disciplina. Uno de los mayores avances de la dialectología del siglo pasado, la dialectometría, se ha convertido en una realidad en prácticamente todas las lenguas cultivadas (Goebel 1992; Nerbonne 2013).

La variedad de técnicas cuantitativas utilizadas en la dialectología pone al alcance de los investigadores un amplio abanico de posibilidades de analizar los datos dialectales. Pero todo análisis cuantitativo necesita de una base de datos amplia que aleja al dialectólogo de las prácticas del denominado (*single*) *feature based* dialectología, ganando en la objetividad de la muestra del análisis.

En este trabajo se presentan los pasos que hay que seguir para desarrollar una investigación en dialectología cuantitativa. Además, se exponen algunas de las técnicas utilizadas, como las destinadas a la cuantificación de la distancia entre variedades, a la clasificación jerárquica, y/o al análisis del *continuum* dialectal. Así mismo, también se exponen métodos multivariantes para la identificación de patrones de variación, estudio de las variables que presentan similares patrones geográficos, analizar la probabilidad de pertenencia a determinados grupos dialectales, etc. La metodología de la dialectología cuantitativa se halla delimitada por los siguientes pasos: elección de un atlas lingüístico del que se proveerá su base de datos (que puede ser fonética, ortográfica o/y etiquetada), aplicación de una medida de distancia que proporciona una matriz de distancias y el uso de técnicas cuantitativas aplicadas a la matriz de distancias. La cuantificación se ha convertido en un paso obligatorio para expertos que se dedican al estudio de la variación lingüística.

Palabras clave: variación geolingüística, dialectometría, metodología.

Abstract

The introduction of the quantification of geolinguistic variation has brought a spectacular rise in publications on the subject, which indicate a renewed vitality of the discipline. One of the greatest advances in dialectology of the last century, dialectometry, has become a reality in practically all cultivated languages (Goebel 1992; Nerbonne 2013).

The variety of quantitative techniques used in dialectometry offers researchers a wide range of possibilities for analyzing dialectal data. But any quantitative analysis needs a broad database that distances the dialectologist from the practices of the so-called '(single) feature based' dialectology, gaining in the objectivity of the analysis sample.

The methodology of quantitative dialectology begins with the choice of a linguistic atlas from which its database will be provided (which can be phonetic, orthographic or/and labeled). The application of a distance measurement provides the distance matrix. The quantitative techniques applied to the distance matrix range from the quantification of the distance between dialectal varieties (interpunctual dialectometry), the hierarchical classification of dialectal varieties, the analysis of the dialectal *continuum* (with the technique of multidimensional scaling (MDS), the analysis of the correlation between geographical and linguistic distance, the detection of linguistic characteristics, etc. Quantification has become a mandatory step for experts who study linguistic variation.

Keywords: Regional variation, dialectometry, methodology.

1. Introducción

El camino recorrido en la cuantificación de la variación geolingüística ha sido muy fructífero desde el punto de vista de los avances, de la variedad de técnicas y del gran número de idiomas en los que se ha aplicado. La dialectometría (en adelante DM), que es como se denomina en este campo a la cuantificación de la variación geolingüística, se encuentra en un momento que era inimaginable cuando se elaboraron los primeros trabajos cuantitativos. Actualmente, se puede decir que la cuantificación es imprescindible en los estudios de la variación geolingüística cuando se trata de clasificar variedades dialectales. Ello no quiere decir que la cuantificación sea el objetivo de los trabajos geolingüísticos, sino la herramienta indispensable para obtener resultados fidedignos, que posteriormente deben de ser interpretados, como en todas las ciencias. Y decimos que es indispensable porque se necesita cuantificar esa variación para trazar y jerarquizar fronteras, delimitar áreas, medir la distancia lingüística entre localidades o variedades dialectales, etc. La DM no es el fin de nada, sino el medio que nos provee de herramientas adecuadas para que el análisis tenga mayor precisión y exactitud.

El gran acierto de Jean Séguy y de los creadores de esta disciplina en la década de los 70 del siglo pasado fue su visión de la realidad. La dialectología no podía evolucionar en aquella época en el plano teórico, porque las herramientas que tenía a mano no daban para más: no se podía avanzar en la delimitación y jerarquización de las fronteras, en la distinción entre zonas de transición y frontera dialectal, en la relación entre la distancia geográfica y la distancia lingüística, o en el análisis de los factores extralingüísticos en la diferenciación geolingüística. Hacían falta nuevas herramientas para analizar toda la información recogida en los atlas lingüísticos.

Hay que recordar, sin embargo, que antes de la aparición de la DM se veía ya una inexcusable necesidad de métodos estadísticos. Ya había propuestas cuantitativas en la literatura científica anterior, algunos loables intentos, también palos de ciego. Entre los precursores de la utilización de métodos cuantitativos en la dialectología se pueden citar entre otros a Davis y MacDavid (1950), Reed y Spicer (1952), Atwood (1955) y Houck (1967). En prácticamente todos ellos se conjuga lo cualitativo (elección de características a analizar por el investigador) y la cuantificación (se pueden ver buenos resúmenes de los trabajos de esa época en Shaw 1974, Schneider 1988 o Kretzschmar 1996).

En este sentido, el abanico del análisis cuantitativo se ha expandido enormemente con la

DM. Hay una gran variedad de técnicas que van introduciéndose y abriendo camino. Incluso ya hay investigadores que dividen la DM en una DM clásica o estándar y otra nueva DM (Pickl y Rumph 2012). Muchas de las técnicas se encuentran compiladas en los programas creados *ad hoc* y otras en paquetes del mercado (Grieve et al. 2011) o en R (www.r-project.org/) (Grieve y Jurgens 2019). A decir verdad, la DM ha dado un gran salto cualitativo y sirve actualmente para todo análisis cuantitativo de la variación geolingüística. Pero no solo es avance en el plano técnico o estadístico, sino que también se ha dado un avance en la concepción de qué es lo que hay que medir. La dimensión lingüística de los dialectómetros ha ido creciendo; tanto que, hoy en día, se plantea claramente que son inviables análisis basados en isoglosas (aunque estas sean cuantitativas) o en características elegidas por el investigador. Ante ello se plantean dos direcciones: una dialectología basada en grandes bases de datos o corpus, y una dialectología basada en sistemas o subsistemas lingüísticos. Esta última dirección, nacida a mediados del siglo XX, ha ido tomando más importancia en generaciones posteriores y ha sido tenida en cuenta por la escuela dialectométrica de Salzburgo y sus seguidores con la taxación o etiquetaje de los datos. La dialectología cuantitativa actual presenta medidas tanto para el análisis de la estructura superficial (nivel fonético), como de la estructura profunda (nivel fonológico) de la lengua. En la primera se encuentra la escuela de Groninga y en la segunda la escuela de Salzburgo.

El trabajo pretende dar una visión general de la DM (los diversos pasos en la cadena dialectométrica, diferentes escuelas, herramientas, etc.) y la aportación de la misma a los estudios de la variación geolingüística. Para ello, se ha estructurado en cinco secciones: en la segunda se explica para qué sirve la dialectometría, en la tercera se especifica cómo se trabaja en ella, exponiendo todos los pasos que hay que dar en un análisis dialectométrico; en la cuarta se analiza la relevancia de estos instrumentos metodológicos y en la quinta se exponen las conclusiones.

2. ¿Para qué sirve la dialectometría?

La DM surgió para superar la crisis epistemológica en la que estaba inmersa la dialectología. En efecto, esta disciplina lingüística se encontraba paralizada, sin mecanismos y técnicas adecuadas para avanzar en el conocimiento de la variación geolingüística, mientras que los atlas lingüísticos languidecían en los empolvados e inmanejables volúmenes. Se lamentaba Séguy de su nula utilización cuando afirmaba que “las ricas colecciones que constituyen los atlas lingüísticos permanecen infrautilizados” (1973a: 3) y Alvar, cuando afirmaba lo siguiente:

Los atlas abruma por la inmensa cantidad de datos que encierran y si no facilitamos su consulta, quedarán como gigantescos archivos no utilizados. Desgraciadamente, tal es la situación en el mundo hispánico (1984: 112-113).

Se puede afirmar que la DM ha dado una segunda vida a los atlas lingüísticos que posibilita su explotación integral. La DM proporciona las herramientas adecuadas de las que carecía la dialectología tradicional para el análisis integral de la inmensa cantidad de datos que albergan los atlas lingüísticos, posibilitando la proyección cartográfica sintética de los mismos. La DM ha abierto una nueva vía a la dialectología, la vía de la cuantificación.

La DM sirve para cuantificar y objetivar los estudios sobre la variación geolingüística; cuantifica la variación, es decir, propone sistemas para contar las diferencias lingüísticas entre localidades y variedades dialectales. Goebel expresó concisamente el tema con la ecuación siguiente: “Dialectometría = geografía lingüística + taxonomía numérica” (1983: 113), que posteriormente desarrolló de la manera siguiente:

Todas estas ciencias [botánica, biología...] se hallan ante el mismo dilema, es decir, la misma necesidad: controlar la desconcertante riqueza de los datos empíricos y clasificarlos en categorías para extraer una visión global, tipológica (1983: 113-114).

3. ¿Cómo se trabaja en DM para el análisis de datos lingüísticos?

Los métodos dialectométricos se han ido poco a poco diversificando y cada día hay más técnicas de análisis cuantitativos y más posibilidades de cartografiar dichos resultados. En sus primeros 50 años de existencia la DM ha hecho un recorrido ingente. Desde la distancia interpuntual de los primeros trabajos (Séguy 1973b; Guiter 1973; Goebel 1976) se pasa a la distribución de similitudes y se llega hasta el estudio de las zonas de transición con lógica *fuzzy*, pasando por la DM dendrográfica o clasificación jerárquica, la correlativa, etc. Repasaremos las principales técnicas utilizadas a lo largo de estos años.

Goebel ha explicado en numerosas ocasiones la mecánica de los trabajos dialectométricos o “cadena dialectométrica” (véase, por ejemplo, Goebel 2010). A continuación se explican en detalle los distintos pasos que se siguen en el desarrollo de esta técnica.

3.1. Tipificación de los datos

El primer paso en la cadena dialectométrica es la decisión de elegir los datos del atlas lingüístico en cuestión para el análisis dialectométrico. Se pueden tomar los datos en su integridad o solamente una muestra de los mismos, atendiendo a criterios objetivos (despojo de mapas que contienen lagunas en las respuestas, etc.). Posteriormente se decide la forma de los datos a analizar; se pueden tomar los datos directamente del atlas lingüístico, sin ningún tipo de tratamiento, o transformados de acuerdo a algún tipo de análisis lingüístico (por ejemplo, una lematización de los datos lexicales, o una tipificación o etiquetaje de los datos gramaticales). Esta fase es mucho más importante de lo que a primera vista pueda parecer, puesto que es el fundamento de la diferenciación lingüística. En el fondo lo que se está decidiendo es si la diferenciación lingüística se va a basar en la estructura superficial o en la estructura subyacente de la lengua. Si los datos del atlas lingüístico pasan directamente a la base de datos sin ningún análisis lingüístico (tanto en alfabeto fonético como en el ortográfico), se estará analizando la estructura superficial de la lengua, puesto que el análisis cuantitativo se llevará a cabo con datos de pronunciación, de la actuación. Sin embargo, si los datos son sometidos a un análisis lingüístico (se puede utilizar cualquier teoría lingüística para ello), la investigación será llevada a cabo de acuerdo a la estructura subyacente de la lengua en cuestión.

3.2. Creación de la base de datos

Una vez decidido el nivel lingüístico de los datos, se ha de crear la base de datos. Esta base de datos puede ser exhaustiva (que contenga todos los datos del atlas lingüístico elegido) o puede contener solo una muestra de los mismos. En este caso el requisito que debe cumplir es el de la representatividad, tanto desde el punto de vista de las características lingüísticas como de la población. Se trata del primer eslabón en la objetividad del análisis: cuanto más se aleje del (*single*) *feature based* dialectología más se acercará a la objetividad en la selección de los datos.

En geolingüística es una práctica habitual el uso de muestras, ante la imposibilidad de analizar el universo lingüístico que albergan las variedades dialectales; de hecho, un atlas lingüístico es una muestra que se supone es representativa del conjunto de rasgos de las variedades analizadas. La representatividad tiene que ser ratificada tanto en el plano lingüístico como en el de las poblaciones.

3.3. Elección de la unidad de distancia lingüística

Sin duda uno de los pasos más importantes dados por la DM ha sido el uso de una unidad de distancia para medir la similitud (o distancia) entre variedades. Es decir, ¿qué y cómo contar las diferencias lingüísticas? ¿En qué basarse para determinar que dos palabras o dos características lingüísticas son iguales o diferentes? Mayoritariamente han sido tres las unidades utilizadas a lo largo de estos años. Séguy (1971: 336; 1973b: 11) utilizó la distancia de Hamming, que consiste en el número de caracteres que difieren de una cadena a otra. Por su parte, Goebel usa una medida nominal o categórica (1983: 115), para ello se hace necesario tipificar las formas recogidas en los atlas lingüísticos. A este proceso de tipificación Goebel lo denomina “taxación” (etiquetaje). Como unidad de medida usa el índice relativo de identidad-IRI (Goebel 1987: 70), renombrado índice relativo de similitud-IRS (Goebel 2013: 145), un índice no ponderado en el que todos los *taxats* o etiquetas tienen idéntico valor (1).

Goebel utiliza también el índice de distancia-IRD. Como estas dos distancias se miden en porcentajes, la $IRD = 100 - IRS$ (si la IRS es igual a 40, la IRD será 60). Tanto la IRS como la IRD son unidades de porcentajes de similitud o distancia que hay entre dos localidades o variedades teniendo en cuenta todas las características que han sido tenidas en cuenta para la investigación. La suma de las similitudes y las disimilitudes o distancias (= diferencias) es igual a 100.

(1)	<i>apatx</i>	<i>azazkal</i>	<i>índice de similitud IRS = 0</i>
			<i>índice de distancia IRD = 100%</i>

Goebel también propuso una unidad de distancia ponderada, denominada índice ponderado de identidad-IPI (Goebel 1987: 74), en el que toman más peso las formas menos usuales y menos expandidas.

La otra unidad de distancia que ha tenido gran recorrido es la distancia o algoritmo de *Levenshtein*, también llamado “distancia de edición” (*edit distance*) o “distancia de cadena” (*string distance*) y que se basa en la eliminación, sustitución o adición de sonidos para pasar de una palabra a otra o de una variante a otra (véase, entre otros, Kessler 1995; Nerbonne y Heeringa 2010; Valls et al. 2012). Esta unidad se usa para

medir la distancia fonética entre dos palabras basándose en la pronunciación, pero tiene el inconveniente de que no es una medida adecuada cuando se trata de contar las distancias entre dos palabras etimológicamente diferentes. No tiene sentido medir, por ejemplo, la distancia fonética entre las palabras vascas *apatx* y *azazkal* que significan ‘pezuña’, puesto que se trata de dos palabras completamente diferentes (2).

(2)	<i>apatx</i>	<i>azazkal</i>	<i>diferencia</i>
	<i>a</i>	<i>a</i>	<i>0</i>
	<i>p</i>	<i>z</i>	<i>1</i>
	<i>a</i>	<i>a</i>	<i>0</i>
	<i>tx</i>	<i>z</i>	<i>1</i>
		<i>k</i>	<i>1</i>
		<i>a</i>	<i>1</i>
		<i>l</i>	<i>1</i>

La diferencia entre estas dos palabras según la distancia *Levenshtein* es 5/7 (0,71). No es una medida adecuada, porque son dos vocablos diferentes y las coincidencias de sonidos son fortuitas.

Aunque no han logrado la popularidad de las anteriores, son conocidas también otras unidades de distancia como: distancia Euclídea, distancia Manhattan, distancia Canberra, distancia binaria o Minkowski, distancia cofenética, distancia distributiva o χ^2 , también llamada CHI-2 y KHI-2 (Philps 1984), entre otras.

Si el uso de una técnica estadística apropiada es importante, tiene mayor importancia si cabe la elección de una buena medida. Para ello se ha de saber qué es lo que queremos medir y cómo lo vamos a hacer. Podemos medir la distancia lingüística de los dialectos en la estructura fonética o superficial, o en la estructura subyacente o fonológica, con análisis lingüístico previo de los datos. Como ya ha quedado demostrado (Clua 2010) los dos tipos de análisis son interesantes, pero hay que tener en cuenta que los resultados son muy diferentes. Es, por tanto, labor del dialectólogo decidir el tipo de datos que debe utilizar en su investigación. De acuerdo con la decisión del plano lingüístico, el dialectólogo ha de decidirse por una u otra medida de distancia.

3.4. Matriz de distancias

La aplicación de la unidad de distancia a la base de datos proporciona la matriz de distancias. Esta matriz reúne las distancias lingüísticas entre todas las poblaciones analizadas. De hecho, se pueden crear matrices de distancias diferentes, según la medida de distancia que se aplique: si se aplica la media de similitud-IRS se obtendrá una matriz de similitudes, en la que el 100 % indica identidad absoluta; en cambio, si se aplican otras medidas de distancia, como la *Levenshtein*, se obtendrán matrices de distancia, en la que 100 % indicará máxima distancia (Figura 1).

	Ahetze	Aia	Aldude	Alkotz	Altzai	Altzurükü	Amezketeta	Andoain	Aniz	Aramaio
Abaurregaina	40,61	43,99	35,64	33,05	48	47,8	46,59	45,74	35,07	53,69
Ahetze		41,85	26,25	36,92	40,65	41,56	43,3	43,68	35,81	52,92
Aia			46,54	37,74	53,22	54,53	22,51	21,97	41,89	44,14
Aldude				37,92	38,59	39,15	46,96	47,08	35,96	52,14
Alkotz					50,74	49,44	37,94	37,96	23,7	48,96
Altzai						14,17	56,12	54,77	49,98	60,03
Altzurükü							56,51	55,77	48,9	60,31
Amezketeta								21,31	40,05	43,52
Andoain									41,65	44,76
Aniz										50,15

Figura 1. Matriz de distancias de localidades vascas con datos del EHHA (parte de la matriz)

Como se puede observar en la Figura 1, las distancias lingüísticas entre localidades se expresan utilizando el mismo sistema de las distancias geográficas expresadas en km. Así, por ejemplo, la distancia lingüística entre la localidad de Abaurregaina y Aia es un 43,99 % y de la misma localidad a Aramaio 53,69 % (es decir, entre las dos localidades más de la mitad de las características analizadas son diferentes).

La matriz de distancias lingüísticas puede ser un punto de partida interesante para tipos de análisis como el grado de dialecticidad que puede albergar una lengua, análisis de las localidades más aisladas, etc.

3.5. Análisis estadísticos de los datos

Una vez lograda la matriz de distancias lingüísticas el dialectólogo tiene ante sí dos puntos a decidir: la técnica estadística a llevar a cabo y la visualización de los resultados. El repertorio de técnicas estadísticas es grande y han sido muchas las utilizadas en la dialectometría. En esta contribución se citarán las más fructíferas en este campo del saber.

3.5.1. Primer paso: distancias entre localidades o variedades dialectales

La proyección de la cuantificación de las similitudes o diferencias lingüísticas entre localidades y variedades dialectales se llevó en un primer momento a mapas, y posteriormente también a diferentes recursos visuales.

La distancia interpuntual fue el primer análisis dialectométrico que se diseñó (Séguy 1973b; Guiter 1973; Goebel 1976). La distancia interpuntual proporciona las distancias lingüísticas entre dos localidades, tanto contiguas como alejadas. El procedimiento habitual es la visualización de la distancia entre localidades colindantes por medio de mapas isoglóticos: mapas en los que cada localidad analizada está representada por un polígono y los lados de los polígonos representan una “isoglosa cuantitativa” que permite visualizar el número de desigualdades entre dos localidades vecinas. Tanto Séguy (en lo que denominó DM lineal o unidireccional), como Guiter (en lo que denominó método global) y Goebel (en lo que denominó dialectometría interpuntual) utilizaron este procedimiento. Se trata de un paso importante en la cuantificación de las distancias lingüísticas entre localidades. De esta manera, Guiter, por ejemplo, pudo establecer una jerarquía de fronteras dialectales (1973: 63) con datos de diversos atlas lingüísticos del sur y sureste de Francia que abarcaban el vasco, el catalán y el occitano.

Otro de los resultados de la proyección de las distancias interpuntuales son los mapas de similitudes, en los que partiendo de una localidad se pueden visualizar las diferencias con las demás localidades. Utilizando esta técnica se pueden crear tantos mapas como

localidades se hayan analizado (Figura 2). Al proyectar los resultados en el mapa, los polígonos de las localidades son coloreados de acuerdo a la similitud que guardan entre ellas, lo que se representa por medio de colores (los colores rojos representan la máxima similitud, los colores azul oscuro, la mínima similitud). Se trata de una técnica cartográfica utilizada por geógrafos y usuarios de análisis de conglomerados jerárquicos, que fue incorporada a la DM por Goebel (1987).

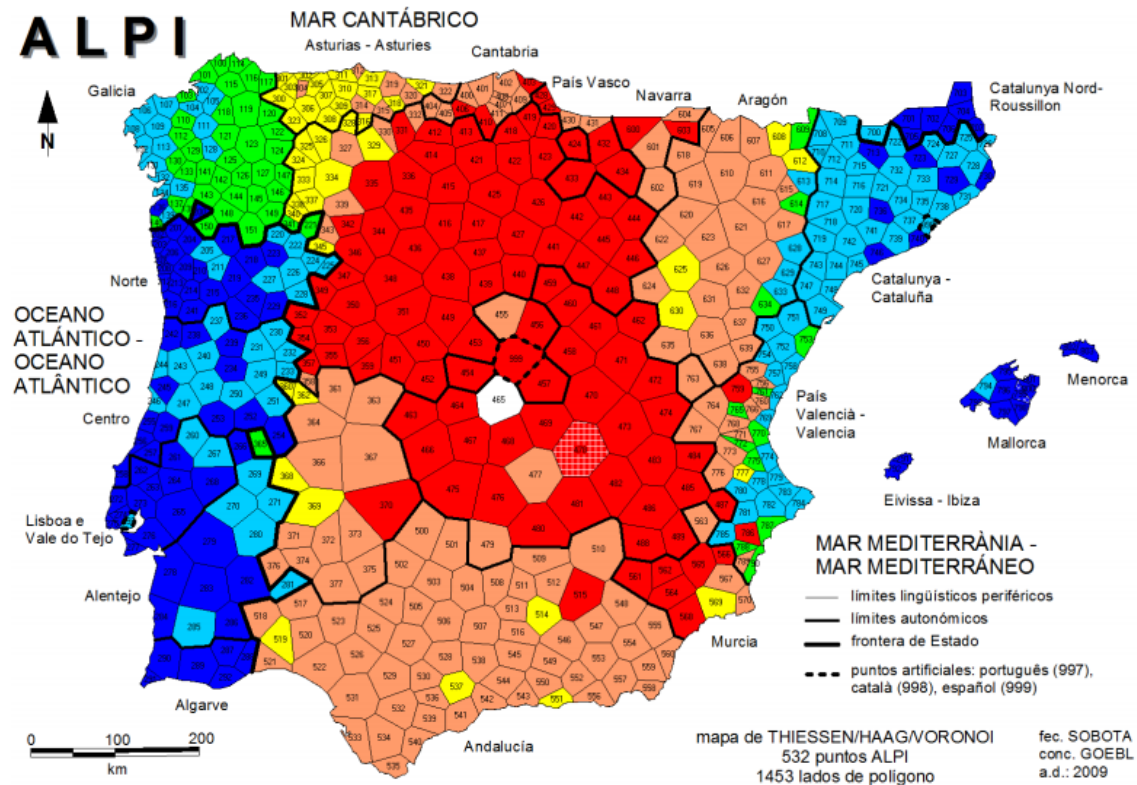


Figura 2. Distribución de similitud de la localidad Camarenilla-465 (Toledo) con relación a las hablas locales (ALPI) (mapa cedido por Goebel)

Las distancias lingüísticas entre localidades pueden ser agrupadas en distintos grupos. En el mapa de la Figura 2, se visualizan en 6 grupos (localidades en rojo, naranja, amarillo, verde, azul claro y azul oscuro). Como se puede comprobar en la Figura 2, la localidad de Camarenilla (polígono blanco en el centro del mapa) tiene altos niveles de similitud con localidades de su entorno y zona norte de la Península (localidades en rojo) y buenos niveles con localidades andaluzas y aragonesas (localidades en naranja). Sin embargo, la similitud decae según se acerca a zonas catalanas, gallegas (amarillos, verdes, azul claros y oscuros) y, sobre todo, portuguesas (con la mayoría de sus localidades en azul oscuro)¹.

3.5.2. Segundo paso: clasificación de los dialectos

Sin embargo, la gran explosión de la dialectología cuantitativa se dará con la aplicación de técnicas aglomerativas multivariantes (*multivariate analysis*). La DM dio un salto espectacular con el empleo de técnicas estadísticas más sofisticadas y de mayor alcance

en la delimitación de las distancias y fronteras dialectales. La posibilidad de la jerarquización de dichas fronteras (fronteras bruscas y fronteras tenues, por ejemplo) y la clasificación de los dialectos vino de la mano del análisis de conglomerados o clasificación aglomerativa jerárquica (*cluster analysis*), llamada también DM dendrográfica y clasificación jerárquica. La dialectometría aglomerativa pone al alcance del investigador todo un arsenal de técnicas (métodos jerárquicos y no jerárquicos) y algoritmos (distancias mínimas, distancias máximas, distancias entre centroides, distancias ponderadas, Ward o varianza mínima, vecino más alejado, o más próximo), con la posibilidad de combinar todos ellos con distintas unidades de distancia.

A día de hoy, la dialectometría aglomerativa se ha convertido en una técnica estándar para la detección y clasificación de variedades y áreas dialectales, pero, al mismo tiempo, se trata de la parte más extraña para el lingüista no habituado en estadística. Por todo ello, se hace necesario aclarar algunos conceptos.

El primero de ellos es el concepto de “algoritmo”, que se puede definir como un sistema determinado de agrupar las variedades basado tanto en la distancia intragrupal (distancia que hay dentro de un grupo de variedades dialectales) o intergrupala (distancia que hay entre dos grupos de variedades dialectales diferentes).

El segundo de los conceptos es el “análisis de conglomerados”. Este análisis consiste en agrupar localidades dentro de una jerarquía, en la que paso a paso se van aglutinando empezando por las dos localidades que tengan más elementos en común; es decir, las dos localidades más próximas lingüísticamente. Una vez agrupadas las dos primeras localidades en adelante funcionan como si fueran una única localidad. Se recurre otra vez al mismo procedimiento para buscar las siguientes localidades más próximas lingüísticamente y se funden en un grupo. El mecanismo se va repitiendo hasta que todas las localidades se hallen agrupadas (Goebel 1991, 1992 y ss.; Nerbone et al. 2008: 2). El principio general del análisis de conglomerados es, por tanto, agrupar un gran conjunto de datos en subconjuntos más pequeños, en el que cada subconjunto sea lo más homogéneo posible en sí mismo y tan distinto como sea posible con los demás subconjuntos. Mediante esta técnica todas las variedades analizadas se integran en una jerarquía (Figura 3).

El gráfico resultante, llamado dendrograma, señala perfectamente la jerarquía en la que se engloba cada variedad dialectal. El dendrograma posibilita jerarquizar las fronteras que van surgiendo según se van agrupando las variedades. Cuanto más alto sea el nivel de configuración de un conglomerado la frontera que dibuja en el mapa tendrá mayor impacto y será lingüísticamente más relevante.

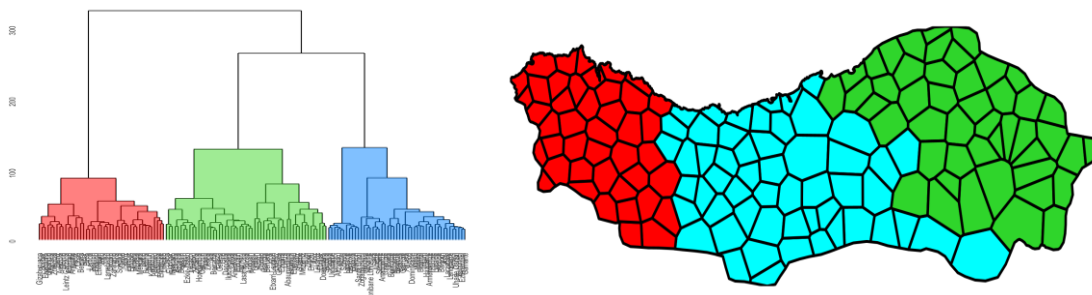


Figura 3. Dendrograma y mapa de los dialectos del euskera, respectivamente

El dendrograma (el gráfico de la izquierda de la Figura 3) recoge todas las variedades analizadas (en la parte baja del dendrograma) en el estudio de la clasificación de los dialectos vascos. El dendrograma, en su parte izquierda, se halla provisto de una escala que agrupa la distancia lingüística que contienen las variedades. En este caso la escala llega hasta el nivel 360, que es la diferenciación total entre todas las variedades; a mayor diferenciación lingüística entre localidades, más alto será el nivel de la escala. Para visualizar los dialectos en el mapa se ha dado un corte en el nivel 160 de la escala del dendrograma; el corte presenta las variedades agrupadas en tres grupos (rojo, verde y azul). A este nivel se halla representada el 44,44 % de la varianza (el nivel 360 representa el 100 % de la varianza). La proyección cartográfica del dendrograma refleja tres dialectos o grupos dialectales: occidental (en rojo), central (azul) y oriental (verde). Hay que señalar, sin embargo, que las fronteras entre la zona roja y la azul, y la frontera entre la zona azul y la verde no se encuentran al mismo nivel de la escala, puesto que el grupo rojo se une al grupo formado por el azul y el verde en el nivel más alto de la escala, mientras que estos dos grupos se unen más tempranamente. La frontera que separa el grupo rojo del azul es jerárquicamente superior a la frontera que separa los grupos azul y verde.

El mapa visualiza el espacio geográfico de los grupos dialectales definidos en el dendrograma. En este caso, los tres grupos dialectales han sido considerados como “dialectos”. Se trata de tres dialectos (el occidental, el central y el oriental) con una distribución geográfica compacta. Es de resaltar que la frontera entre el dialecto occidental y el central es idéntica a la frontera trazada por otros procedimientos de la dialectología tradicional.

Los resultados obtenidos mediante esta técnica cuantitativa son mucho más fiables que las clasificaciones dialectales logradas por métodos de la dialectología tradicional. A pesar de ello, hay que decir que no se trata de una técnica perfecta. Estas clasificaciones pueden sufrir cierta inestabilidad ya que pequeños aportes de datos pueden derivar en diferencias significativas en los resultados. Por esta razón, es necesario que los resultados sean corroborados con otras técnicas. Nerbonne et al. (2008) proponen dos técnicas para superar esta inestabilidad: por un lado, el *bootstrapping* (denominado también *bootstrap clustering*) o método de remuestreo y, por otro, el *noisy clustering* o agrupamiento con ruido.

En primer lugar, el *bootstrapping* o método de remuestreo es una técnica de análisis que lleva a cabo diversos muestreos con reemplazamiento de datos en cada uno de ellos. El proceso de agrupación se repite un determinado número de veces con elección de diversos elementos de muestra en cada uno de ellos. El resultado final es un dendrograma compuesto (*composite dendrogram*) en el que consta el número de veces que aparece el grupo en las distintas iteraciones (véase, por ejemplo, Nerbonne et al. 2008: 651).

Por su parte, el *noisy clustering* o agrupamiento con ruido funciona a partir del cálculo de la varianza en la matriz de distancias, así mismo, se especifica un techo de ruido y se repite la operación 100 veces. El resto del procedimiento es idéntico al *bootstrapping*. Pero ¿qué es el ruido en una matriz de datos? El ruido en los datos está relacionado con la estabilidad y la variabilidad desde el punto de vista estadístico; cuando hay mucha inestabilidad y gran variabilidad se dice que hay mucho ruido en los datos y suele estar relacionado con errores tipográficos en la entrada de datos, formas raramente

documentadas, variables exógenas, etc. (véase Wieling y Montemagni 2016).

En todos los casos, la visualización de los resultados se logra por medio del dendrograma y de la cartografía. El dendrograma fruto de la clasificación ha de ser interpretado en primer lugar y, posteriormente, se ha de elegir una técnica de validación de la mejor partición para su proyección al mapa. Esta labor es facultad del investigador, pero la estadística proporciona una serie de criterios (coeficiente de correlación cofenético, coeficiente de pertenencia, replicación, simulaciones Monte Carlo, interpretabilidad teórico-práctica...) para certificar que dicha partición es correcta (véase más en Prokić y Nerbonne 2008; Mucha y Haimerl 2005).

3.5.3. Tercer paso: el análisis del *continuum* dialectal

La técnica aglomerativa divide las áreas dialectales de forma rígida, en grupos compartimentados en los que las fronteras dialectales se dibujan abruptas. Sin embargo, no siempre se da este tipo de fronteras entre áreas dialectales. Esta técnica no refleja siempre de modo adecuado la realidad dialectal, porque no siempre se da una frontera clara. Esta técnica anterior, por tanto, no visualiza los casos de *continuum* dialectal, en el que el paso de un área a otra se hace de forma gradual. La visualización de este *continuum* dialectal ha sido posible mediante la técnica *Multidimensional Scaling* (MDS) o escalamiento multidimensional, que fue aplicada por primera vez en dialectometría por Embleton (1987a, 1987b), siendo desarrollada más tarde por Embleton et al. (2013) y Nerbonne (2010), entre otros.

El escalamiento multidimensional reduce las dimensiones de la base de datos a una, dos o tres dimensiones. Las variedades analizadas se ubican en un espacio multidimensional abstracto. Esta técnica presenta las similitudes lingüísticas en un espacio bi o tridimensional, en el que la distancia lingüística entre variedades y áreas dialectales se expresa en el espacio con mayor o menor proximidad entre ellas, sin necesidad de ninguna frontera. Se han desarrollado diversas opciones para visualizar los resultados: 2D diagrama, 2D diagrama con localidades conectadas por líneas con el centro del área, 3D diagramas e incluso con posibilidad de interactuar para seleccionar el mejor plano (Embleton et al. 2013: 15), diagramas en diversos colores y mapas (véase Figura 4).

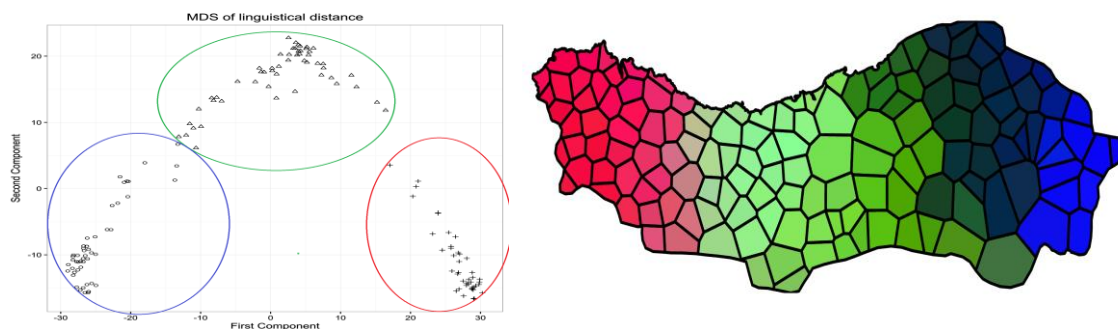


Figura 4. Diagrama y mapa correspondiente al MDS con datos del euskera

El gráfico de la Figura 4 presenta las variedades del euskera en una V invertida. Se han analizado los datos en tres dimensiones que en el gráfico se señalan con símbolos y en el mapa con colores. El gráfico señala claramente un alejamiento lingüístico entre las variedades que se hallan en la parte derecha (en círculo rojo, que corresponden a las

variedades en rojo en el mapa), mientras que no hay una clara separación entre las otras dos áreas dialectales (círculos en verde y azul) que en el mapa se señalan en color verde y azul. Esto indica que la frontera entre el área verde y azul no es tan nítida y abrupta como la frontera entre el área roja y verde.

3.5.4. Cuarto paso: el estudio de la relación entre la distancia geográfica y la distancia lingüística

La influencia de la geografía en la variación lingüística ha sido un tema recurrente en la historia de la geolingüística desde sus comienzos. La creencia general es que la distancia lingüística crece a medida que se va alejando desde un punto de partida. Esta relación no ha podido ser concretada hasta que los dialectólogos han podido disponer de herramientas cuantitativas. Fue Séguy el primero que formuló, en su curva de correlación (1971: 348), los principios de la DM correlativa, con datos de 16 atlas lingüísticos. Posteriormente, el tema ha sido una y otra vez analizado por distintos investigadores y distintos datos: Goebel con datos del *Atlas linguistique de la France-ALF* (2005) y *Atlas Lingüístico de la Península Ibérica-ALPI* (2013) (véase Figura 5), Nerbonne, en solitario y con colaboradores (2010, 2013; Nerbonne y Heeringa 2007; Spruit et al. 2008; Nerbonne et al. 2010, entre otros), Aurrekoetxea con datos del corpus *Bourciez* (2010, 2016), etc. Esta relación ha hecho avanzar los estudios sobre los factores externos de la variación geolingüística hasta tal punto que Nerbonne ha propuesto un *Principio Fundamental de la Dialectología*, según el cual “las variedades lingüísticas próximas son generalmente –pero no siempre– más similares que las lejanas” (Nerbonne 2013: 222).

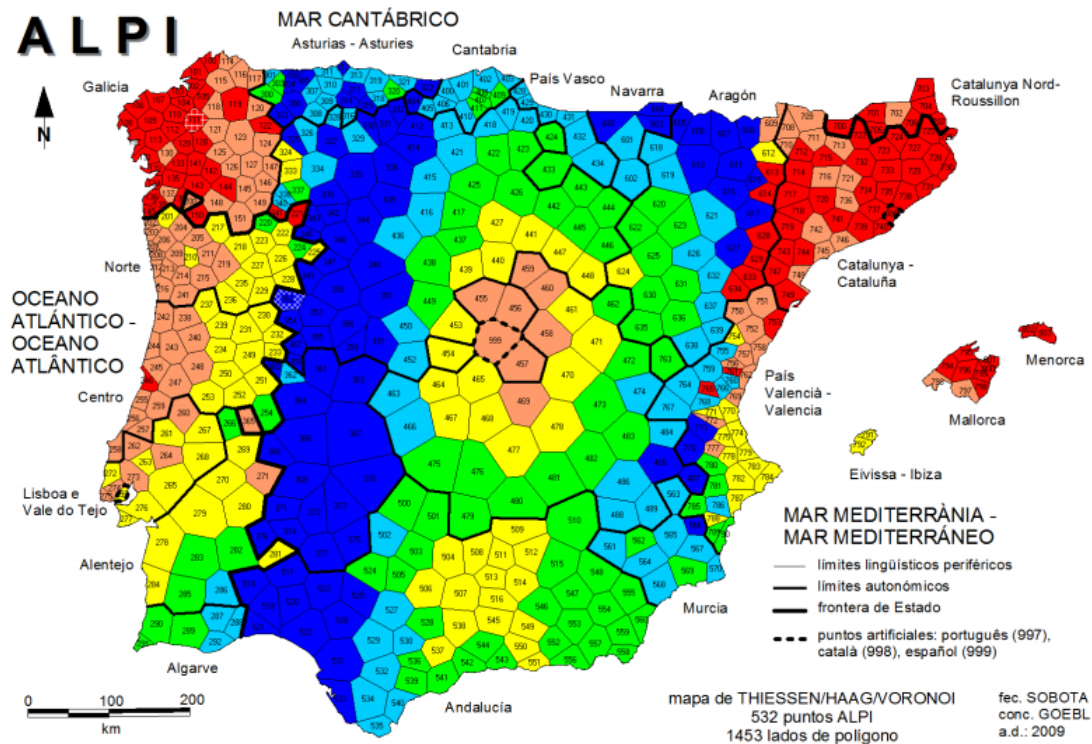


Figura 5. Correlación entre la distancia geográfica y la distancia lingüística en ALPI (mapa cedido por H. Goebel)

La Figura 5 presenta la correlación entre la distancia geográfica y la distancia lingüística en la Península Ibérica. Como se puede observar, hay una alta correlación en Galicia, Cataluña, Portugal y la zona central (zonas en rojo, naranja y amarillo); sin embargo, hay una baja correlación en regiones marcadas en azul, tanto oscuro como claro, y verde.

El factor de la distancia geográfica como el factor determinante en la diferenciación lingüística ha sido, sin embargo, puesto en duda: así, por ejemplo, Szmrecsanyi (2012) sostiene que la distancia geográfica es un mal predictor de la variabilidad morfosintáctica con datos del inglés. Otros investigadores propugnan el estudio de la correlación entre el tiempo necesario para trasladarse de una localidad a otra con la distancia lingüística. Por ejemplo, Gooskens (2005) analiza el tiempo de trayecto como indicador de la distancia lingüística: cuanto más tiempo de trayecto para trasladarse de una localidad a otra, mayor diferenciación lingüística hay entre dichas variedades. Posteriores investigaciones avalan esta vía.

Aparte de esta correlación la dialectología cuantitativa ha llegado a analizar la correlación entre distintas categorías gramaticales; como, por ejemplo, la correlación entre la distancia fonológica y morfológica, o entre esta última y la léxica, etc. De hecho, se pueden analizar las correlaciones entre todas estas categorías gramaticales.

3.6. Otros métodos dialectométricos

Si bien tanto el análisis de conglomerados como el escalamiento multidimensional han sido los métodos más utilizados en la dialectometría clásica, se han utilizado y siguen siendo utilizados cantidad de métodos y técnicas cuantitativas. En este apartado se expondrán sucintamente algunos de ellos: *Dual Scaling* o análisis de correspondencias, auto-correlación espacial, análisis de factores, análisis de componentes principales, agrupamiento difuso, glotograma, agrupamiento de bipartitos espectrales y dialectometría inversa.

3.6.1. *Dual Scaling* o análisis de correspondencias

Cichocki (1989) fue el primero en implementar esta técnica en geolingüística con el estudio sobre las frecuencias de tres variables en el francés canadiense. Los resultados de este análisis se representan en dos espacios multidimensionales, un espacio de localidades y un espacio lingüístico. En el diagrama la proximidad entre localidades indica afinidad lingüística, de tal modo que localidades lingüísticamente afines se encuentran más cercanas que las localidades lingüísticamente más alejadas. El espacio de las variantes lingüísticas se superpone a las localidades. La técnica del *dual scaling* extrae los patrones lingüísticos más importantes de las variedades y los representa en dos gráficos. Es una técnica que ayuda a conocer la estructura geolingüística de ciertas variables lingüísticas, pero no sirve para el estudio de muchas variables a la vez.

3.6.2. Auto-correlación espacial

Esta técnica fue propuesta por Kretzschmar (1992, 1996) y desarrollada por Grieve et al. (2011). En ella las variables lingüísticas se someten a un análisis de auto-correlación espacial para identificar patrones significativos de variación lingüística regional, que es

similar a proyección de isoglosas. Se trata de una medida de dependencia espacial que permite medir el grado de agrupamiento espacial en los valores de una variable.

3.6.3. Análisis de factores

El análisis de factores ha sido utilizado en más de una ocasión: Chauveau (1985) fue el primero en utilizarlo. Posteriormente Aurrekoetxea (1995), Inoue (2004), Shackleton (2005), Nerbonne (2006) y Leinonen (2008), entre otros, lo han utilizado en sus investigaciones. Se trata de una técnica de reducción de datos que permite al investigador descubrir las variables que muestran patrones similares de variación y si las variables se pueden dividir en subconjuntos independientes. Las variables que se correlacionan entre sí se combinan en factores, lo que significa que la cantidad total de datos se puede reducir.

3.6.4. *Principal Components Analysis*-PCA o análisis de componentes principales

El análisis de componentes principales es un método multivariante para capturar la variabilidad intrínseca en los datos: una combinación de características, las cuales explican la varianza en los datos de la mejor manera posible. Este es el primer componente principal. Posteriormente se procede buscando repetidamente una combinación de características que no esté correlacionada con las del componente principal o anteriores y que explica la mayor parte de la variación restante en los datos. Los resultados pueden ser cartografiados: las localidades encuadradas en cada componente principal son ploteadas con distintos colores (Shackleton 2005; Hyvönen et al. 2007; Ueda 2017).

3.6.5. *Fuzzy clustering* o agrupamiento difuso

El *fuzzy clustering* es un tipo de análisis de conglomerados que analiza la probabilidad de un agrupamiento de las variables analizadas. Hay algunas aplicaciones en geolingüística como Meschenmoser y Pröll (2012), Pröll (2013) o Aurrekoetxea, Iglesias et al. (2020) para el análisis de zonas de transición.

En la DM clásica la información detallada contenida en los mapas individuales se pierde en el proceso de construcción de áreas o grupos dialectales; es decir, se puede saber cómo se distribuyen los dialectos, pero no hay información sobre cuáles son las variables que actúan en cada caso. Se utiliza este procedimiento para construir grupos de mapas identificables que comparten patrones espaciales similares de variantes. Este método tiene como objetivo preservar la información contenida en mapas individuales y mantenerla accesible en cada paso en lugar de agregarlo, de esta forma no hay pérdida de información contenida en las variantes. Desde este punto de vista, el enfoque depende de la calidad de los métodos de búsqueda de corpus para recuperar características y compararlas con otras. De esta forma se crean grupos de mapas que exhiben similares propiedades estructurales. La similitud de dos mapas se mide mediante la función de covarianza.

La covarianza describe la relación de dos variables aleatorias y muestra valores positivos altos si las cantidades tienden a comportarse de manera similar, esto es, valores altos de A implican valores altos de B y valores bajos de A implican valores

bajos de B. La covarianza toma valores negativos en caso contrario, es decir, cuando estas implicaciones no se cumplen y valores altos de A no implican valores altos de B. La covarianza no depende de la posición de las cantidades de interés sino solo de su distancia. De esta manera, la función de covarianza describe la estructura de las propiedades de los mapas. Como la similitud se basa en la función de covarianza, dos mapas son similares si comparten las mismas propiedades estructurales. Estas varianzas, por tanto, son analizadas por *fuzzy cluster* análisis.

De este modo, esta técnica estadística proporciona la oportunidad de tratar los mapas de los atlas lingüísticos como un corpus, accesibles a través de procesos automatizados. Ello proporciona una nueva posibilidad de estudio de los corpus dialectales, que hasta ahora estaba restringida al uso de datos categóricos, así como a la falta de detección de patrones espaciales.

3.6.6. Glotograma

Esta técnica combina la edad de los locutores con el factor geográfico y parte de la premisa de que el lenguaje se difunde geográfica y socialmente. El glotograma fue concebido en la geolingüística japonesa en la década de 1980 y se utiliza para estudiar la dinámica de las lenguas o variedades en contacto. Los resultados se recogen en mapas y gráficos con signos en los que se señala el cambio lingüístico que se produce en el parámetro temporal y espacial, todo ello de acuerdo a la edad de los locutores y sus localizaciones (Sanada 2010).

3.6.7. *Bipartite-Spectral graph* o agrupación de gráficos bipartitos espectrales

La DM aglomerativa ha sido criticada porque ha priorizado el estudio de áreas dialectales en detrimento de las variables lingüísticas. En los últimos años se está revirtiendo la situación. Wieling y Nerbonne (2009, 2011) propusieron una nueva técnica que busca grupos de variedades y simultáneamente la asociación de sonidos relacionados con esos grupos. Por consiguiente, con esta técnica se estudia la relación dialectal de las localidades y su base lingüística conjuntamente. Para ello se miden las distancias de las localidades con el estándar usando la distancia *Levenshtein* (véase sección 3.3.). El objetivo no es lograr cuáles son los sonidos más característicos de cada grupo o los sonidos que más veces aparecen en dichas localidades. En realidad, el objetivo es, por una parte, lograr grupos de localidades con respecto a la distancia con la variedad estándar y, por otra, la correspondencia de sonidos entre la variedad estándar y los grupos de localidades.

Una variante es el *Hierarchical Spectral Partitioning of biparte graph* (Wieling y Nerbonne 2010) que conlleva una ventaja importante sobre otros métodos dialectométricos, puesto que la base lingüística se determina simultáneamente, cerrando la brecha entre la dialectología tradicional y la cuantitativa. Además de mostrar que los resultados del agrupamiento jerárquico mejoran con respecto al método de agrupamiento espectral plano utilizado en un estudio anterior (Wieling y Nerbonne 2009), esta técnica se utiliza para identificar las correspondencias de sonidos más importantes para cada grupo. Esta es una ventaja importante con respecto al método jerárquico, puesto que evita la necesidad de métodos externos para determinar las correspondencias de sonido más importantes para un grupo geográfico.

3.6.8. *Reverse dialectometry* o dialectometría inversa

La dialectometría inversa propone analizar la distribución espacial de cada una de las características lingüísticas. Se trata de detectar e identificar variables gramaticales midiendo la afinidad entre ellas, en función de las áreas que marca cada variable. Este punto de vista ha sido denominado como *reverse dialectometry* (Craenenbroeck 2014) o *variant-based dialectometry* (Pickl y Rumpf 2012). Aurrekoetxea et al. (2021) indagan en la misma vía, analizando las variables con un clúster jerárquico y un MDS de las características verbales de la lengua vasca, y llegan a la conclusión de que algunas características lingüísticas tienen los mismos patrones geolingüísticos.

3.7. Programas integrales creados para la DM

Los paquetes estadísticos del mercado presentan inconvenientes para su aplicación en la geolingüística. Entre estos se pueden citar, por ejemplo, que estos paquetes no presentan ningún tipo de cartografía, lo que hace prácticamente inviable su uso, o que no tienen en cuenta las múltiples respuestas. Por todo ello, los programas integrales de dialectometría son una buena opción, puesto que presentan únicamente las técnicas estadísticas requeridas en la DM y suelen incluir la opción de trasladar a un mapa los resultados de los procesos estadísticos. Actualmente los dialectómetros tienen distintas herramientas hechas exclusivamente para el estudio cuantitativo de la variación geolingüística. En las siguientes líneas se glosarán algunos de las más accesibles y eficaces. Es de destacar que todos ellos son programas de libre acceso.

3.7.1. *Visual Dialectometry-VDM* (Universidad de Salzburgo)

La primera herramienta de DM, denominada *Visual Dialectometry-VDM* (<http://ald.sbg.ac.at/dm>), fue diseñada por H. Goebel, creada por Haimerl y desarrollada en la Universidad de Salzburgo². Goebel ha publicado en más de una ocasión las características de esta herramienta (Goebel 2006; Goebel y Smečka 2016).

Se trata de un programa que funciona en local, en la plataforma *Windows*, con dos unidades de distancia categóricas (índice relativo de similitud-IRS e índice ponderado de similitud-IPS) y una base de datos numeral (el dialectólogo debe adaptar sus datos lingüísticos a numéricos). Por todo ello, ha tenido un gran impacto en la geolingüística, sobre todo románica. El mayor hándicap de la herramienta es su programa cartográfico, que debe implementarse imperativamente en Salzburgo.

Mediante esta herramienta se pueden lograr mapas de similitud, mapas isoglóticos (*honeycomb map / isoglotic maps*), mapas de rayos (*beam map*), mapas de zonas de transición (desviación típica), mapas de zonas conservadoras (*Skewness o coefficient d'asymétrie de Fisher*), mapas de zonas de mayor cohesión lingüística (máxima, *synopsis of the maxima of N similarity distributions*), DM correlativa (correlación entre la distancia geográfica y la diferencia lingüística o entre dos categorías gramaticales), análisis de conglomerados (con diversos algoritmos: *Complete, Ward, Average...*), algoritmos de visualización / discretización de las distancias lingüísticas (MedMinMax, Med, MedMW), etc.

3.7.2. *Gabmap* (Escuela de Groninga)

Gabmap es una aplicación web destinada a facilitar las exploraciones en dialectología cuantitativa asistida por computadora (<https://www.gabmap.nl/>). *Gabmap* crea varias vistas de datos de dialectos, desde histogramas de caracteres utilizados para detectar errores de codificación, pasando por alineaciones de transcripciones fonéticas utilizadas para medir la distancia de pronunciación, hasta diagramas de escala de colores multidimensionales destinados a ilustrar resultados cuantitativos de forma intuitiva, etc. Muchos análisis van acompañados de instalaciones que permiten a los investigadores estudiar más, y por ejemplo, buscar las bases lingüísticas más importantes de una división de área, o examinar los resultados de la agrupación para la confiabilidad estadística.

El software, escrito de código fuente abierto (Nerbonne et al. 2011), funciona tanto con datos categóricos (lexicales o gramaticales), como con datos numéricos (vectores de frecuencias de formantes de las vocales...). Está dirigido especialmente al análisis de datos fonéticos y al estudio de las distancias de pronunciación entre localidades. Para ello usa la unidad de distancia *Levenshtein*. El usuario tiene, además, a su disposición distintas técnicas de análisis como mapas de rayos, análisis de conglomerados, MDS, agrupamiento difuso, agrupamiento con ruido, etc.

3.7.3. *Diatech* (UPV/EHU)

La herramienta *Diatech* es también una aplicación web (<http://eudia.ehu.es/diatech/>), libre y multilingüe (Aurrekoetxea et al. 2013; Aurrekoetxea et al. 2016). Uno de los motivos principales para la creación de la herramienta fue posibilitar el tratamiento de las respuestas múltiples (MR) o polimorfismo (Aurrekoetxea, Nerbonne, Rubio 2020). En los paquetes estadísticos convencionales tan solo se podía introducir un dato por cada casilla de la base de datos. Por consiguiente, si se tienen datos con MR y se quieren analizar todos los datos a la vez, sin ningún tipo de selección y sin pérdida de información, inevitablemente se ha de elegir un programa *ad hoc*, como el *Diatech*. Por ejemplo, la herramienta permite la introducción de más de una respuesta a cada pregunta en la misma localidad. Un botón de referencia: El mapa correspondiente al caso “motivativo indeterminado” de la declinación de las palabras terminadas en la vocal “o” en vasco (*Euskararen Herri Hizkeren Atlasa-EHHA V*: mapa 1082); de 145 localidades, se han recogido más de una respuesta en 56 localidades.

La aplicación funciona con datos en alfabeto normativo, fonético o lematizado / etiquetado. Además, pone a disposición del usuario tres medidas de distancia: índice de similitud, índice ponderado de similitud y la distancia *Levenshtein*; es decir, unidades de distancia utilizadas tanto en la escuela de Salzburgo como Groninga. Los análisis estadísticos que se pueden hacer con esta herramienta son los siguientes: mapas sinópticos (mapas de similitudes, áreas de transición, correlación entre distancias geográficas y lingüísticas...), mapas de rayos, mapas isoglóticos, análisis de *cluster*, MDS y agrupamiento difuso.

3.7.4. *ProDis* (UB)

La DM o el análisis cuantitativo de los datos dialectales se ha expandido a todas las categorías lingüísticas. Quizás la que más ha tardado en ser analizada cuantitativamente

ha sido la prosodia. Con este fin, es decir, el de analizar los datos prosódicos con datos numéricos, el grupo del Laboratorio de Fonética de la UB (Elvira-García et al. 2018; Fernández Planas et al. 2019) implementaron un programa para el tratamiento cuantitativo de los datos acústicos llamado *ProDis*. El programa provee similitudes acústicas entre localidades, clasificaciones jerárquicas y agrupamientos multidimensionales.

3.8. Otras vías para hacer DM

El avance en dialectología cuantitativa ha sido inmenso, inimaginable al comienzo de su andadura (una buena visión de las técnicas utilizadas en *aggregate variationist analyses* se puede consultar en Nerbonne y Wieling 2018). En los últimos años ha ido creciendo la implementación de técnicas cuantitativas a la variación geolingüística. El *software R* ha propiciado una nueva vía para ello. La proliferación de técnicas no hace más que enriquecer el estudio en este campo. La DM se ha convertido, por tanto, en una tarea interdisciplinaria en el que la colaboración entre dialectólogos, geógrafos, matemáticos e informáticos es esencial. Una muestra de esta interdisciplinaria se puede encontrar en Jeszenszky et al. (2018), en estudios basados en sistemas complejos con datos recogidos de Twitter (Gonçalves y Sánchez 2016; Donoso y Sánchez 2017; Ruiz-Tinoco 2018), en el uso de *USTM autoencoder* (Rama y Çöltekin 2016) o en técnicas para reducir los cuestionarios representativos mediante Coeficiente Simple de Emparejamiento y métodos *k-means* (Aurrekoetxea, Clua et al. 2020), entre otros.

4. ¿Por qué es relevante este instrumento metodológico?

La relevancia de la aportación de la DM a los estudios sobre la variación geolingüística se puede resumir en cuatro aspectos: (i) el uso de técnicas cuantitativas para la medición de la diversidad lingüística, (ii) la objetividad en la elección de los datos, (iii) la clasificación jerárquica de los dialectos y (iv) los avances recientes de la dialectología.

4.1. Técnicas cuantitativas para la medición de las distancias lingüísticas

El gran paso de Jean Séguy (1973a) fue su decisión de dar cuenta de las diferencias lingüísticas entre las localidades basándose en los datos del *Atlas linguistique de la Gascogne-ALG*. El resultado de esta decisión fue la plasmación de las diferencias lingüísticas entre las localidades en un mapa con la creación de lo que más tarde se denominó isoglosa cuantitativa; es decir, la distancia lingüística entre localidades se visualiza con el grosor de las isoglosas.

Su trabajo manual pasó rápidamente a ser elaborado por procedimientos automatizados y la cuantificación computarizada. Este planteamiento rompe radicalmente con el modo tradicional de investigación en geografía lingüística e inicia un modo de investigación nuevo, desconocido totalmente, y revolucionario con el uso de la estadística. A pesar de que algunos todavía tengan reparos para aceptarlo, lo cierto es que la dialectología se está poniendo en *modo ciencia*. Como ciencia empírica que es, la dialectología comienza a usar lo que se conoce como “método científico”. En concreto se empiezan a utilizar dos procedimientos científicos: la medición y el razonamiento. Y dos principios: el de la reproductibilidad y el de la refutabilidad.

De hecho, esta tarea implica como primer paso la búsqueda de una unidad de medida de las diferencias lingüísticas (véase sección 3.3.). Por esta razón, Séguy, cuando propuso el uso de la '*distancia de Hamming*' como unidad de medida, dio un paso gigante: por primera vez en la historia de la geolingüística se proponía una unidad para medir las diferencias lingüísticas entre localidades. Este hecho tiene una relevancia incuestionable y marca un hito en el desarrollo de la cuantificación de la variación geolingüística. La relevancia no está en que haya acertado con una unidad de medida determinada, sino en el hecho mismo de proponer una unidad de medida.

En definitiva, la DM fundamentalmente proporciona una vasta colección de herramientas y técnicas para el análisis de la variación geolingüística, de tal forma que actualmente no hay dificultad alguna para medir cualquier distancia lingüística, ya sea fonética (segmental o suprasegmental), fonológica, sintáctica, morfológica o léxica. Por consiguiente, posibilita cuantificar las diferencias lingüísticas entre localidades, jerarquizar las fronteras dialectales, clasificar jerárquicamente áreas dialectales, detectar zonas conservadoras y áreas de transición, visualizar la correlación entre la diferenciación lingüística y la distancia espacial o geográfica, el influjo de la densidad lingüística en la variación, la visualización del *continuum* dialectal, etc.

4.2. Objetividad

La objetividad que proporciona la DM se basa en dos pilares: (i) el uso de grandes masas de datos y (ii) el uso de técnicas estadísticas de análisis.

Uno de los puntos más débiles de la dialectología tradicional era el uso de datos elegidos por el investigador, que, aunque no fuera de manera consciente, estaban sujetos a la subjetividad del autor. Por esta razón, el uso de amplias bases de datos, creadas con criterios objetivos, ha sido otra de las aportaciones de la DM a los estudios de la geografía lingüística. Gracias a este paso, se consiguió un cambio en el paradigma de estudio: de los (*single*) *feature-based Variation Studies*, estudios basados en unas pocas características (Nerbonne 2009) a la *aggregate dialectology* o dialectología aglutinativa. Además, esta decisión conlleva el uso exhaustivo de los atlas lingüísticos. Séguy expuso claramente esta determinación:

Proponemos determinar la diferencia dialectal basándonos no en unos pocos criterios elegidos arbitrariamente, sino en la integración de todos los datos contenidos en los seis volúmenes del *Atlas linguistique de la Gascogne* (1973a: 21).

Si bien los primeros proyectos de atlas lingüísticos no implicaban la creación de bases de datos, pronto comenzaron los proyectos atlantográficos: para su creación, se recurrió a la informática, y para su proyección, a la cartografía (se recomienda consultar Baiwir y Renders 2013 para la discusión sobre si los atlas lingüísticos pueden ser considerados como corpus). No hay que olvidar que los primeros dialectómetros tuvieron que crear sus bases de datos partiendo de datos de atlas lingüísticos ya publicados (por ejemplo, Séguy con su *Atlas linguistique de la Gascogne-ALG* y Goebel con su proyecto de dialectometrización del *Atlante Italo-Svizzero-AIS*, *Atlas linguistique de la France-ALF* o *Atlas linguistique de la Península Ibérica-ALPI*, entre otros). Todo el tiempo, el esfuerzo humano y económico invertido en estas tareas ha merecido la pena, sin duda

alguna. Sin un atlas lingüístico o proyecto semejante de recogida y ordenación de los datos sería y es impensable hablar de DM. Los atlas lingüísticos son un requisito básico.

Cuando no es posible utilizar los atlas lingüísticos en su integridad, el dialectólogo debe atenerse a los requisitos estadísticos de las muestras, el segundo principio de objetividad anteriormente expuesto. La fiabilidad de la muestra es una condición básica en toda investigación puesto que la objetividad es uno de los factores de la fiabilidad. En lingüística, como en otras ciencias, es necesario utilizar muestras debido a la imposibilidad de abarcar la lengua o la variedad de una región en su conjunto y en su totalidad. De esta manera, el objetivo de los lingüistas es que su muestra de datos sea representativa del universo que pretende representar. En este sentido, la inmensa mayoría de los atlas lingüísticos son muestras representativas, sobre todo, en el aspecto léxico, morfológico y fonético. Quizás el aspecto sintáctico es el que queda infrarrepresentado, especialmente en los atlas más clásicos.

4.3. Clasificación dialectal jerárquica

La DM ha contribuido a solucionar el problema de la clasificación dialectal. No hay duda de que el análisis de conglomerados ha dado consistencia y seguridad a las clasificaciones dialectales. Las clasificaciones previas, llevadas a cabo por métodos tradicionales, han de ser revisadas y contrarrestadas con las realizadas con técnicas estadísticas. A pesar de que todo análisis estadístico reduce los factores de variabilidad, resaltando los factores que más incidencia tienen en ella, y que suponen una simplificación de la realidad lingüística, la clasificación obtenida por estos métodos cuantitativos es mucho más fiable que la llevada a cabo por métodos manuales.

El análisis de conglomerados, por otra parte, puede ser interpretado tanto desde el punto de vista sincrónico como del diacrónico: Goebel (2002) ha dado muestras de ello con datos del *Atlas linguistique de la France-ALF*. La interpretación sincrónica del dendrograma se apoya en una lectura ascendente en la creación del dendrograma y visualiza la diferenciación lingüística actual entre grupos dialectales, mientras que la interpretación diacrónica, en una lectura descendente, explica las fragmentaciones sucesivas a lo largo de la historia de dichas variedades lingüísticas.

4.4. Avance de la dialectología

La DM ha posibilitado un avance teórico de la dialectología. Algunas cuestiones que hasta ahora no podían ser resueltas o no podían avanzar han tenido un mayor recorrido gracias a la aportación de la DM. Por ejemplo, desde los primeros trabajos dialectológicos, nadie discutía que había una relación estrecha entre la distancia geográfica y la distancia lingüística. Sin embargo, esa suposición no ha podido ser formulada cuantitativamente hasta que los trabajos dialectométricos han determinado porcentajes de correlación que varían de una lengua a otra, de un contexto lingüístico a otro. A consecuencia de esto, se sabe que el factor de distancia geográfica no es un factor predominante en la variación geolingüística, o, al menos, no en todos los casos. Así mismo, gracias a los avances dialectométricos, hoy en día, se está en disposición de jerarquizar las clasificaciones de las variedades dialectales con la suficiente seguridad.

Bien es cierto que quedan por resolver muchos problemas de la geolingüística, sobre todo, en la interpretación de los resultados estadísticos. Por ejemplo, el dialectómetro ha

sido incapaz hasta el momento de denominar las áreas creadas por él mismo. Se da a este respecto una paradoja interesante: la dialectología tradicional, cuando llega a definir áreas dialectales, no duda ni un instante en nombrar dichas áreas y no tiene ningún reparo en clasificar dichas zonas como dialectos, subdialectos, variedades... Hay muy pocos casos en los que dudan o incluso prefieran no utilizar la denominación “dialecto”, y uno de ellos es el gallego. García de Diego (1959: 130; 1984: 155) habla de la no existencia de dialectos en el gallego por ser poco profundas las diferencias entre las distintas modalidades, Zamora Vicente (1953: 80) habla de dos subdialectos. Fernández Rei corrobora la idea de García de Diego de que “non se pode falar de dialectos propiamente ditos no galego” (1990: 36), idea que ha sido corroborada por Dubert-García (2020) con procedimientos estadísticos.

5. Conclusiones

En este estudio, por consiguiente, se han expuesto detenidamente los pasos a seguir en la tarea dialectométrica, desde la elección del atlas lingüístico, la construcción y el etiquetaje de la base de datos, la elección de la unidad de distancia lingüística pasando por la elección de las distintas técnicas cuantitativas en consonancia a los objetivos de la investigación.

Como conclusión general, por una parte, hay que destacar que la DM ha delimitado la distancia lingüística en términos cuantitativos y ha posibilitado el avance teórico en temas propios de la dialectología, como la distancia entre variedades, la clasificación de los dialectos, la relación entre la distancia geográfica y la distancia lingüística, etc. Por otra parte, se han creado y desarrollado distintas técnicas utilizadas en DM para medir las distancias lingüísticas, delimitar grupos dialectales (MDS y el análisis de *cluster*), y delimitar y analizar las zonas de transición y zonas arcaicas, etc. Pero, ante todo, la aportación de la DM ha sido clave en el uso de grandes masas de datos para eliminar definitivamente la subjetividad del dialectólogo en la selección de las características. La superación de la fase *Single feature-based dialectology* para llegar al *aggregate Dialectology* ha sido un paso crucial.

En definitiva, la relevancia de la DM se mide en las técnicas cuantitativas para la medición de las distancias lingüísticas, en la objetividad de los resultados, en las técnicas jerárquicas de clasificación de variedades y en posibilitar el avance teórico de la dialectología en general.

6. Referencias

- Alvar, Manuel. 1984. Automatización de los índices en los atlas lingüísticos. En M. Alvar, ed. *Informática y Lingüística*. Málaga: Ágora, pp. 107-119.
- Atwood, E. Bagby. 1955. The phonological divisions of Belgo-Romance. *Orbis* IV: 367-389.
- Aurrekoetxea, Gotzon. 1995. *Bizkaieraren egituraketa geolinguistikoa*. Bilbao: UPV/EHU.
- Aurrekoetxea, Gotzon. 2010. The correlation between morphological, syntactic and phonological variation in the Basque language. En B. Heselwood y Cl. Upton,

- eds. *Proceedings of Methods XIII. Papers from the Thirteenth International Conference on Methods in Dialectology, 2008*. Frankfurt: Peter Lang, pp. 207-218.
- Aurrekoetxea, Gotzon. 2016. Distantzia geografikoaren eta hizkuntza distantziaren arteko korrelazioa. En G. Aurrekoetxea, J. M. Makazaga y P. Salaberri, eds. *Txipi Ormaetxea omenduz. Hire bordatxoan*, Bilbao: UPV/EHU, pp. 55-72.
- Aurrekoetxea, Gotzon; Clua, Esteve; Iglesias, Aitor; Usobiaga, Iker; Salicrú, Miquel. 2020. Characterizing dialect groups: distance and informativeness associated with linguistic forms. *Zeitschrift für Dialektologie und Linguistik* 2020.2: 307-326. DOI: 10.25162/zdl-2020-0011
- Aurrekoetxea, Gotzon; Fernández-Aguirre, Karmele; Rubio, Jesús; Ruiz, Borja; Sánchez, Jon. 2013. 'DiaTech': A New Tool for Dialectology. *Literary and Linguistic Computing* 28.1: 23-30. DOI: 10.1093/lc/fqs049
- Aurrekoetxea, Gotzon; Iglesias, Aitor; Clua, Esteve; Usobiaga, Iker; Salicrú, Miquel. 2020. Analysis of transitional areas in Dialectology: Approach with Fuzzy Logic. *Journal of Quantitative Linguistics*. DOI: 10.1080/09296174.2020.1732765
- Aurrekoetxea, Gotzon; Iglesias, Aitor; Santander, Gotzon; Usobiaga, Iker. 2016. Diatech: tool for making dialectometry easier. *Dialectologia* 17: 1-22.
- Aurrekoetxea, Gotzon; Abasolo, Juan; Clua, Esteve; Usobiaga, Iker; Salicrú, Miquel. (En prensa). Hizkuntza-aldagaien aldakortasuna: lehen hurbilketa. *Uztaro* 118, 61-79.
- Aurrekoetxea, Gotzon; Nerbonne, John; Rubio, Jesús. 2020. Unifying analyses of multiple responses. *Dialectologia* 25: 59-86.
- Baiwir, Esther; Renders, Pascale. 2013. Les atlas linguistiques sont-ils des corpus? *Corpus* 12: 27-37.
- Chauveau, Jean-Pierre. 1985. Etude de cartes linguistiques par l'analyse factorielle des correspondances. En A. Moll, eds. *Actes del XVI CILFR* (Ciutat de Mallorca, 7-12 d'abril de 1980), 2. Palma de Mallorca, pp. 379-397.
- Cichocki, Wladyslaw. 1989. An application of dual scaling in Dialectometry. *Journal of English Linguistics* 22.1: 91-95.
- Clua, Esteve. 2010. Relevancia del análisis lingüístico en el tratamiento cuantitativo de la variación dialectal. En G. Aurrekoetxea y J. L. Ormaetxea, eds. *Tools for Linguistic Variation*. Bilbao: UPV/EHU, pp. 151-166.
- Craenenbroeck, Jeroen van. 2014. Quantity and quality in linguistics-reverse dialectometry. En *Methods in Dialectology XV. August 11–15 4, 2014*, Groningen, the Netherlands.
<https://static1.squarespace.com/static/5217e223e4b090faa01f8f2d/t/55a4e0eee4b0d12431c376a2/1436868846371/project-phd.pdf>
- Davis, Alva L.; MacDavid, Raven I. 1950. A transition area. *Language* 26.2: 264-273.
- Donoso, Gonzalo; Sánchez, David. 2017. Dialectometric analysis of language variation in Twitter. En P. Nakov, M. Zampieri, N. Ljubešić, J. Tiedemann, S. Malmasi y A. Ali, eds. *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects*. Association for Computational Linguistics, pp. 16-25. DOI: 10.18653/v1/W17-12
- Dubert-García, Francisco. 2020. O galego, unha lingua sen dialectos: olladas sociais e lingüísticas sobre a variación dialectal. *Estudios Románicos* 29: 147-163.
- EHHA: Euskaltzaindia. 2013. *Euskararen Herri Hizkeren Atlas*. Vol. V. Bilbao: Euskaltzaindia.

- Elvira-García, Wendy; Balocco, Simone; Roseano, Paolo; Fernández-Planas, Ana María. 2018. ProDis: A dialectometric tool for acoustic prosodic data. *Speech Communication* 97: 9-18.
- Embleton, Sheila. 1987a. A new technique for dialectometry. En V. B. Makkai, ed. *Twelfth LACUS Forum*. Jupiter Press: Lake Bluff, Illinois.
- Embleton, Sheila. 1987b. Multidimensional scaling as a dialectometrical technique. En R. M. Babitch, ed. *Papers from the Eleventh Annual Meeting of the Atlantic Provinces Linguistic Association*. New Brunswick: Centre Universitaire de Shippagan in Shippagan, pp. 33-49.
- Embleton, Sheila; Uritescu, Dorin; Wheeler, Eric S. 2013. Defining dialect regions with interpretations: Advancing the multidimensional scaling approach. *Literary and Linguistic Computing* 28.1: 13-22.
- Fernández Planas, Ana María; Elvira-García, Wendy; Balocco, Simone; Roseano, Paolo. 2019. Análisis dialectométrico con ProDis: un paso más en los estudios prosódicos de AMPER. En J. Dorta, ed. *Investigación geoprosódica. Amper: análisis y retos*. Madrid/Frankfurt am Main: Iberoamericana Vervuert, pp. 119-135.
- Fernández Rei, Francisco. 1990. *Dialectoloxía da lingua galega*. Vigo: Xerais.
- García de Diego, Vicente. 1990. *Elementos de gramática histórica gallega (fonética-morfología)*. Burgos: Hijos de Santiago Rodríguez.
- García de Diego, Vicente. 1959. *Manual de dialectología española*. Madrid: Ediciones Cultura Hispánica.
- García de Diego, Vicente. 1984. *Elementos de gramática histórica gallega (Fonética-Morfología)*. Burgos, 1906.
- Goebel, Hans. 1976. La dialectométrie appliquée à l'ALF (Normandie). En A. Várvaro, ed. *XIV Congresso internazionale di linguistica e filologia romanza. Atti., II*. Napoles/Amsterdam: G. Macchiaroli / J. Benjamins, pp. 165-195.
- Goebel, Hans. 1983. Matrices, distances et interpoints. Etude dialectométrique sur les isoglosses quantitatives. En Marie-Rose Simoni-Aurembou, ed. *Cahier des Annales de Normandie n°15. Dialectologie et littérature du domaine d'oïl occidental: actes du colloque tenu à l'Université de Caen en février 1981* pp. 113-137. DOI: 10.3406/annor.1983.3899
- Goebel, Hans. 1987. Points chauds de l'analyse dialectométrique: pondération et visualisation. *Revue de linguistique romane* 51: 63-118.
- Goebel, Hans. 1991. Una classificazione gerarchica di dati geolinguistici tratti dall'AIS. Saggio di dialettometria dendrografica. *Linguistica* 31:341-351.
- Goebel, Hans, 1992. Problèmes et méthodes de la dialectométrie actuelle (avec application á l'ais). En G. Aurrekoetxea y X. Videgain, eds. *Actas del congreso internacional de dialectología*. Bilbao: Euskaltzaindia, pp. 429-475.
- Goebel, Hans. 2002. Analyse dialectométrique des structures de profondeur de l'ALF. *Revue de linguistique romane* 66: 5-63.
- Goebel, Hans. 2005. La dialectométrie corrélative: un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme. *Revue de linguistique romane* 69: 321-367.
- Goebel, Hans. 2006. Recent advances in Salzburg Dialectometry. *Literary and Linguistic Computing* 21.4: 411-435.
- Goebel, Hans. 2010. Introducción a los problemas y métodos según los principios de la Escuela Dialectométrica del Salzburgo (con ejemplos sacados del "Atlante italo-

- Svizzero” AIS). En G. Aurrekoetxea y J. L. Ormaetxea, eds. *Tools for linguistic variation*. Bilbao: UPV/EHU. Anejos del Seminario de Filología Vasca Julio Urquijo LIII, pp. 3-39.
- Goebel, Hans. 2013. La dialectometrización del ALPI: rápida presentación de los resultados. En E. Casanova Herrero y C. Calvo Rigual, eds. *Actas del XXVI Congreso Internacional de Lingüística y de Filología Románicas (Valencia 2010)*. Berlin/Boston: Walter de Gruyter, vol. VI, pp. 143-154.
- Goebel, Hans; Smečka, Pavel. 2016. The quantitative nature of working maps (WM) and taxatorial areas (TA). A brief look at two basic units of Salzburg Dialectometry (S-DM). *Studies in Quantitative Linguistics* 23. *Issues in Quantitative Linguistics* 4: 11-127.
- Gonçalves, Bruno; Sánchez, David. 2016. Learning about Spanish dialects through Twitter. *RILI* XVI-2: 65-75.
- Gooskens, Charlotte. 2005. Traveling time as a predictor of linguistic distance. *Dialectologia et Geolinguistica* 13: 38-62.
- Grieve, Jack; Jurgens, David. 2019. Mapping word frequencies on Twitter using R and Python. En *Workshop presented at NWA 48, University of Oregon, 2019-10-10*. http://jurgens.people.si.umich.edu/tutorials/Mapping_Word_Frequencies_on_Twitter_using_Python.html
- Grieve, Jack; Speelman, Dirk; Geeraerts, Dirk. 2011. A statistical method for the identification and aggregation of regional linguistic variation. *Language Variation and Change* 23: 1-29.
- Guiter, Henri. 1973. Atlas et frontières linguistiques. En G. Straka y P. Gardette, eds. *Les dialectes de France à la lumière des atlas régionaux (Colloque de Strasbourg, 1971)*. Paris: CNRS, pp. 61-109.
- Houck, Charles L. 1967. A computerized statistical methodology for linguistic geography: a pilot study. *Folia linguistica* 1.1-2: 80-95.
- Hyvönen, Saara; Leino, Antti; Salmenkivi, Marko. 2007. Multivariate analysis of Finnish dialect data – An overview of lexical variation. *Literary and Linguistic Computing* 22.3: 271-290.
- Inoue, Fumio. 2004. Multivariate analysis, geographical gravity centers and the history of the standard Japanese forms. *Area and Culture Studies* 68: 15-36.
- Jeszszky, Péter; Stoeckle, Philipp, Glaser, Elvira; Weibel, Robert. 2018. A gradient perspective on modeling interdialectal transitions. *Journal of linguistic geography* 6.2: 78-99.
- Kessler, Brett. 1995. Computational dialectology in Irish Gaelic. En *Proceeding of the European ACL*. Dublin: ACL, 60-67.
- Kretschmar, William. 1992. Isoglosses and predictive modeling. *American Speech* 67: 227-249.
- Kretschmar, William A. 1996. Quantitative areal analysis of dialect features. *Language Variation and Change* 8: 13-39.
- Leinonen, Therese. 2008. Factor analysis of vowel pronunciation in Swedish dialects. *International Journal of Humanities and Arts Computing* 2.1-2: 189-204.
- Meschenmoser, Daniel; Pröll, Simon. 2012. Using fuzzy clustering to reveal recurring spatial patterns in corpora of dialect maps. *International Journal of Corpus Linguistics* 17: 176-197.
- Mucha, Hans Joachim; Haimerl, Edgar. 2005. Automatic validation of hierarchical cluster analysis with application in Dialectometry. En C. Weihs y W. Gaul, eds.

- Classification—the ubiquitous challenge. Proc. of 28th Mtg Gesellschaft für Klassifikation, Dortmund, Mar. 9-11, 2004.* Berlin: Springer, pp. 513-520.
- Nerbonne, John. 2010. Mapping aggregate variation. En A. Lameli, R. Kehrein y S. Rabanus, eds. *Language and space. International handbook of linguistic variation. Vol. 2.* Berlin: De Gruyter Mouton, pp. 476-495.
- Nerbonne, John; Colen, Rinke; Gooskens, Charlotte.; Kleiweg, Peter; Leinonen, Therese. 2011. Gabmap – a web application for Dialectology. *Dialectologia* Special issue II: 65-89.
- Nerbonne, John; Heeringa, Wilbert. 2007. The geographic distribution of linguistic variation. *Studies in generative grammar*: 267-298. <https://www.lingexp.uni-tuebingen.de/sfb441/LingEvid2006/abstracts/nerbonne.pdf>
- Nerbonne, John; Heeringa, Wilbert. 2010. Measuring dialect differences. En P. Auer y J. E. Schmidt. eds. *Theories and Methods*, Vol.1. Berlin, New York: De Gruyter Mouton, pp. 550-567.
- Nerbonne, John; Kleiweg, Peter; Manni, Franz; Heeringa, Wilbert. 2008. Projecting dialect distances to Geography: Bootstrap clustering vs. noisy clustering. En Ch. Preisach, H. Burkhardt, L. Schmidt-Thieme y R. Decker, eds. *Data analysis, machine learning and applications. Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation e.V., Albert-Ludwigs-Universität Freiburg, March 7–9, 2007.* Berlin / Heidelberg: Springer, pp. 647-654.
- Nerbonne, John. 2006. Identifying linguistic structure in aggregate comparison. *Literary and Linguistic Computing* 21.4: 463-475.
- Nerbonne, John. 2009. Data-Driven Dialectology. *Language and Linguistics Compass* 3.1: 175-198. DOI: 10.1111/j.1749-818X.2008.00114.x
- Nerbonne, John. 2013. How much does Geography influence language variation? En P. Auer, M. Hilpert, A. Stukenbrock y B. Szmrecsanyi, eds. *Space in Language and Linguistics. Geographical, Interactional, and Cognitive Perspectives.* Berlin: De Gruyter, pp. 220-236.
- Nerbonne, John; Prokić, Jelena; Wieling; Martijn; Gooskens, Charlotte. 2010. Some further dialectometrical steps. En G. Aurrekoetxea y J. L. Ormaetxea, eds. *Tools for Linguistic Variation.* Bilbo: UPV/EHU, pp. 41-56.
- Nerbonne, John; Wieling, Martin. 2018. Statistics for aggregate variationist analyses. En Ch. Boberg, J. Nerbonne y D. Watt, eds. *The handbook of Dialectology.* Oxford: John Wiley & Sons, pp. 400-414.
- Philps, Dennis. 1984. Dialectométrie automatique. En H. Goebel, ed. *Dialectology*, Série Quantitative Linguistics 21. Bochum: Dr. Brockmeyer, pp. 275-296.
- Pickl, Simon; Rumpf, Jonas. 2012. Dialectometric concepts of space: Towards a variant-based dialectometry. En S. Hansen, Ch. Schwarz, Ph. Stoeckle y T. Streck, eds. *Dialectological and folk dialectological concepts of space. Current methods and perspectives in Sociolinguistic research on dialect change*, Series: *linguae & litterae*, 17, Berlin/Boston: De Gruyter, pp. 199-214. DOI: 10.1515/9783110229127
- Prokić, Jelena; Nerbonne, John. 2008. Recognizing groups among dialects. *International Journal of Humanities and Arts Computing* 2.1-2: 153-172.
- Pröll, Simon. 2013. Detecting structures in linguistic maps—Fuzzy clustering for pattern recognition in geostatistical dialectometry. *Literary and Linguistic Computing* 28.1: 108-118.
- Rama, Taraka y Çöltekin, Çağrı. 2016. LSTM autoencoders for dialect analysis.

- Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects*, Osaka, Japan: The COLING 2016 Organizing Committee, pp. 25-32.
- Reed, David W.; Spicer, John L. 1952. Correlation methods of comparing idiolects in a transition area. *Language* 28: 348-359.
- Ruiz-Tinoco, Antonio. 2018. Geocorpus del español de las redes sociales y cartografía automática. *Monográficos SINOELE* 17: 598-608.
- Sanada, Shinji. 2010. The “glottogram”: A geolinguistic tool developed in Japan. *Dialectologia*. Special issue I: 185-196.
- Schneider, Edgar W. 1988. Qualitative vs. quantitative methods of area delimitation in dialectology: A comparison based on lexical data from Georgia and Alabama. *Journal of English Linguistics* 21: 175-212.
- Séguy, Jean. 1971. La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane* 35: 335-357.
- Séguy, Jean. 1973a. La dialectométrie dans l'atlas linguistique de la Gascogne. *Revue de Linguistique Romane* 37: 1-24.
- Séguy, Jean. 1973b. *Atlas linguistique de la Gascogne. Complément du volume VI*. Paris: CNRS.
- Shackleton, Robert G. Jr. 2005. English-American speech relationships: A quantitative approach. *Journal of English Linguistics* 33: 99-160.
- Shaw, David, 1974. Statistical analysis of dialectal boundaries. *Computers and the Humanities* 8.3: 173-177.
- Spruit, Marco René; Heeringa, Wilbert; Nerbonne, John. 2008. Associations among linguistic levels. *Lingua, special issue on Syntactic databases. Selected papers presented in the special session Comparing Aggregate Syntaxes, Digital Humanities conference, Paris, July 6, 2006*: 65-90.
- Szmrecsanyi Benedikt. 2012. Geography is overrated. En S. Hansen, C. Schwarz, P. Stoeckle y T. Streck, eds. *Dialectological and folk dialectological concepts of space*. Berlin: de Gruyter, Walter GmbH & Co., pp. 215-231.
- Ueda, Hiroto. 2017. Two statistical treatments of Spanish vocabulary: Composite indices of frequency and dispersion and principal component analysis applied to ordinal frequencies. *Dialectología*, Special Issue VII: 187-227.
- Valls, Esteve; Nerbonne, John; Prokić, Jelena; Wieling, Martin; Clua, Esteve; Lloret, M-Rosa. 2012. Applying the Levenshtein distance to Catalan dialects: A brief comparison of two dialectometric approaches. *Verba* 39: 35-61.
- Wieling, Martijn; Nerbonne, John. 2009. Bipartite spectral graph partitioning to co-cluster varieties and sound correspondences in dialectology. En M. Choudhury, S. Hassan, A. Mukherjee y S. Muresan, eds. *Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing*. Suntec, Singapore: Association for Computational Linguistics, pp. 26-34.
- Wieling, Martijn; Nerbonne, John. 2010. Hierarchical spectral partitioning of bipartite graphs to cluster dialects and identify distinguishing features. En *Proceedings of the 2010 Workshop on Graph-based Methods for Natural Language Processing, ACL, Uppsala, Sweden, July 16, 2010*, Stroudsborg: Association for Computational Linguistics, pp. 33-41.
- Wieling, Martijn; Nerbonne, John. 2011. Bipartite spectral partitioning for clustering dialect varieties and detecting their linguistic features. *Computer Speech and Language* 25.3: 700-715.
- Wieling, Martijn; Montemagni, Simonetta. 2016. Infrequent forms: Noise or not? En

M.-H. Côté, R. Knooihuizen, J. Nerbonne, eds. *The future of dialects: Selected papers from methods in Dialectology XV*. Berlin: Language Science press, pp. 215-223.

Zamora Vicente, Alonso. 1953. De geografía dialectal: -ao, -an en gallego. *Nueva revista de filología hispánica* VII. 1-2: 73-80.

¹ Los mapas se pueden consultar en la página web: <http://dialektkarten.ch/dmviewer/alpi/index.es.html>

² Para una revisión de los trabajos dialectométricos de la escuela de Salzburgo, se puede consultar la siguiente página web: <http://dialektkarten.ch/dmviewer/index.es.html>

«*Sin documentos no hay historia*»

Aprovechamiento del corpus de documentos aragoneses en lengua catalana

María Teresa Moret Oliver – Javier Giralt Latorre

Universidad de Zaragoza – Academia Aragonesa de la Lengua –

Instituto Universitario en Patrimonio y Humanidades

mmoret@unizar.es / jgiralt@unizar.es

Resumen

Tanto los textos literarios como los no literarios han sido imprescindibles para el estudio de las lenguas desde una perspectiva diacrónica, pero también como una fuente de información para analizar su variación lingüística interna. Si nos centramos en la realidad lingüística de Aragón, comprobaremos que, a diferencia del castellano o el aragonés, no existen obras literarias medievales escritas en catalán, a pesar de la amplia producción en esta lengua en otros territorios de la antigua Corona de Aragón. Por tanto, para conocer el catalán en Aragón de época medieval y moderna, el filólogo se ve obligado a examinar los textos notariales que se redactaron en la Franja de Aragón entre los siglos XIV, XV y XVI, los cuales se han convertido en instrumentos imprescindibles para el estudio de la historia de la lengua catalana en tierras aragonesas. En este trabajo daremos cuenta de la metodología que utilizamos para la localización y el estudio de un corpus documental redactado por profesionales, así como de las destrezas que son necesarias para enfrentarse a un análisis de estas características.

Palabras clave: documentación, historia de la lengua, lenguas de Aragón, catalán en Aragón.

Abstract

Both literary and non-literary texts have been essential for the study of languages from a diachronic perspective, but also as a source of information to analyse the internal linguistic variation of languages. If we focus on the linguistic reality of Aragon, we will see that, unlike Castilian or Aragonese, there are no medieval literary works written in Catalan, despite the extensive production in this language in other territories of the former Crown of Aragon. Therefore, to learn about Catalan in Aragon in medieval and modern times, the philologist is obliged to study the notarial texts that were written in the Aragonese Strip between the 14th, 15th, and 16th centuries, which have become essential tools for the study of the history of the Catalan language in Aragonese lands. In this paper we will give an account of the methodology required to locate and study a corpus of documents written by professionals, as well as the skills needed to carry out an analysis of these characteristics.

Keywords: documentation, history of the language, languages of Aragon, Catalan in Aragon.

1. Introducción

Seguramente sea una obviedad comenzar este trabajo insistiendo en el hecho de que, para describir el momento pretérito de una lengua, debemos recurrir obligatoriamente a testimonios escritos del pasado. Esta constatación, no obstante, emana de uno de los problemas básicos de los estudios históricos, la dificultad de formar un corpus adecuado de estudio que dé cuenta de los objetivos que se persiguen o, dicho de otra forma, de la necesidad de obtener un conjunto de materiales que sean útiles a la hora de analizar un aspecto concreto de la evolución diacrónica, a partir de los cuales se puedan establecer con cierta fiabilidad las características de una lengua en una sincronía concreta.

En las últimas décadas, sin embargo –y a pesar de que la dicotomía entre sincronía y diacronía parecía estar más que superada–, se ha prestado menos atención a los estudios de cariz histórico, con un claro predominio entre los lingüistas sobre lo que se ha considerado *modernidad* frente a *tradicionalismo*. Pero resulta innegable que una buena descripción sincrónica sin más debería sustentarse en una base histórica, que puede proporcionarnos un soporte excepcional cuando se trata de investigar las causas concretas de un fenómeno lingüístico.

Los recursos que se han diseñado para esta cuestión básica han ido variando a lo largo de la historia de la investigación filológica y, en general, su evolución está estrechamente relacionada con la dificultad que supone saber aquello que se pretende conocer a través de los estudios diacrónicos. De hecho, en un primer momento, esta disciplina se concibió como mera herramienta complementaria de la literatura, puesto que lo que pretendía era precisamente indagar en profundidad sobre los textos literarios. Por esta razón, los primeros estudios de historia de la lengua se elaboraron como análisis de las características de la lengua literaria, que perseguían distinguir lo que suponía un rasgo de estilo de aquello que formaba parte de la norma de la época.

Este valor subordinado que se daba a esta disciplina se disipa con el tiempo y, posteriormente, comienza a interesar por sí misma; sobre todo en el momento en el que los estudios históricos se aúnan con la dialectología y comienzan a ofrecer datos sobre los procesos de variación que presentan las lenguas en distintos espacios geográficos. No obstante, y pese a que su evolución es más que evidente en las últimas décadas, si nos fijamos en el caso del catalán, los corpus más completos disponibles hasta el momento todavía se componen, mayoritariamente, de obras literarias.

Por esta razón, para los primeros estudios sobre la lengua catalana del pasado se tomaron como base las distintas crónicas de los reyes, como el *Llibre dels fets del rei Jaume I* –que, a pesar de su carácter historiográfico, tienen un claro componente literario–, o el *Curial e Güelfa*, novela de caballería humanística del siglo XV. De hecho, es tanta la producción literaria del catalán medieval, que ha sido posible desarrollar proyectos como el CICA (*Corpus Informatizat del Català Antic*: <http://cica.cat/>), cuyo corpus está basado principalmente en libros de corte, poesía, prosa de ficción y de caballerías, entre otros. Aunque es cierto que también se incluyen en este corpus algunos libros jurídicos; pero son pocos, o ninguno, los textos de tipo utilitario.

La más que evidente distancia comunicativa que existe en obras como las citadas, obliga necesariamente a concluir que los datos que nos ofrecen, pese a su indudable valor, no muestran en ningún caso más que un determinado uso de lengua; esto es, un idiolecto muy concreto que está, además, fuertemente mediatizado por su ya mencionada

finalidad estética (sin dejar de perder de vista que la escritura siempre presentará un estado parcial e incompleto de la lengua hablada). En esta misma línea, conviene recordar que, dado el peso de ciertas convenciones que afectan a los textos literarios, muchos fenómenos habituales en la lengua no tienen cabida en la literatura, que opta en general por una lengua más elaborada y por los usos más prestigiosos.

En el caso de Aragón –y a diferencia de lo que sucede en la vecina Cataluña, en el Reino de Valencia y en Baleares, donde se desarrolló una importante producción que constituyó una auténtica época dorada para la literatura en catalán–, no existe producción literaria de autores oriundos del área catalanohablante aragonesa de época medieval, salvo las traducciones del s. XIV de Guillem Nicolau.



Figura 1. Mapa de la Franja de Aragón (adaptado de Giralt 2017: 513)

A causa de esta ausencia de obras literarias aragonesas de época medieval redactadas en catalán, es lógico afirmar que las fuentes no literarias son un elemento capital, o más bien imprescindible, para conocer el pasado de dicha lengua en esa región, de modo que nos pueden ofrecer los datos que se extraen de textos más elaborados en otros territorios (obras literarias del resto del ámbito catalán) y corroborar que aquello que hallamos es algo general en el habla y no una licencia poética o un rasgo propio del escritor en cuestión.

Así pues, teniendo en cuenta las condiciones que acabamos de esgrimir, no debería resultar baladí recordar que los archivos históricos constituyen el lugar en el que el filólogo interesado en la evolución histórica del catalán en Aragón (véase Figura 1) va a tener que *localizar* los materiales sobre los que basará sus estudios, puesto que los manuscritos antiguos son la única, o al menos, la más importante fuente de información para el conocimiento de la realidad lingüística de épocas pretéritas. Para ello, además, deberá desarrollar determinadas destrezas prácticas, como son la capacidad de *entender*

la *grafía* de los manuscritos, la *descripción* y *clasificación* de los documentos, la comprensión de la *lengua* en la que están redactados, la ubicación en una *fecha* aproximada (según las características materiales y lingüísticas), y la *edición del texto* con unos criterios coherentes.

En nuestra contribución vamos a detenernos, en primer lugar, en la documentación de archivo, para comentar qué tipos de archivos existen, cuál es la tipología documental que podemos encontrar en ellos y cómo era la escritura de aquellos profesionales de la Edad Media. Y, en segundo lugar, abordaremos cómo debe desarrollarse la tarea de edición de los manuscritos antiguos y qué aspectos lingüísticos pueden resultar interesantes en su análisis (con ejemplos concretos en los niveles gráfico-fonético, morfológico, sintáctico y léxico).

2. La documentación de archivo en la lingüística histórica

En los últimos años, tanto los archivos estatales del Ministerio de Cultura como los autonómicos, han realizado un gran esfuerzo con el objetivo de difundir parte del rico patrimonio documental medieval, pero son todavía pocos los proyectos de distinta índole que se ocupen de la documentación conservada en otro tipo de archivos, sobre todo los ubicados en poblaciones más pequeñas y cuyos fondos no siempre tienen un carácter unitario, como más adelante veremos.

No es ninguna novedad, por tanto, que los filólogos hayan descubierto desde hace ya mucho tiempo la importancia que tienen los archivos históricos y su documentación en los estudios de historia de la lengua, como fuente de información de primer orden para estos trabajos; de este modo, no es de extrañar que se haya advertido un empleo claramente significativo de los materiales que allí se custodian en la investigación diacrónica, muy por encima del uso que se ha hecho, y se hace, de otras fuentes empleadas tradicionalmente en el análisis¹. A este respecto, cabe señalar que son numerosas las ventajas que los materiales de archivo presentan a la hora de estudiar la evolución diacrónica de una lengua; entre ellas, hay que destacar cuatro que tienen una especial incidencia en estas investigaciones, particularmente, desde la perspectiva de la dialectología histórica:

(i) Su carácter de pieza única y original, es decir, de escrito que, por su finalidad puramente administrativa, no haya sido necesario darlo a conocer a un gran público y que, por tanto, no se haya reproducido repetidamente a lo largo del tiempo. Para el filólogo, este factor supone que los problemas en torno a la autenticidad del manuscrito se disipen.

(ii) Su precisa ubicación espacial y temporal, puesto que generalmente estos materiales presentan la fecha de redacción y el lugar donde se escribieron, lo que les confiere una especial relevancia para describir un estado de lengua en una diatopía y una sincronía concreta (véase Figura 2 para un caso ilustrativo).

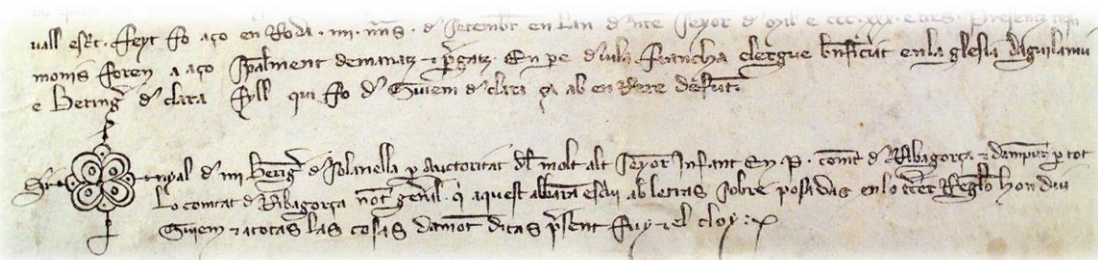


Figura 2. Pergamino 175 del Fondo de «Roda de Isábena» (ACL) fechado el 2 de septiembre de 1333. Not. Berenguer de Solanella: «Feyt fo açò en Roda, IIII nonas de setembre, en l'an de Nostre Senyor de mil e CCC XXX^a e tres. [...] Sinyal de mi Beringer de Solanella, per auctoritat del molt alt senyor infant en Pere, comte de Ribagorça et d'Ampúries, per tot lo comtat de Ribagorça notari general [...]».

De hecho, no son extraños los errores producidos en estudios diacrónicos que han usado copias, sobre todo a la hora de datar determinadas voces o fenómenos lingüísticos. Por todo ello, un investigador debe conocer todas aquellas cuestiones que le permitan establecer que un tipo de letra pertenece a una época determinada y no a otra, qué rasgos lingüísticos se dan en un periodo o en otro (fonéticos, gráficos, morfológicos y sintácticos) y si esos fenómenos que localiza en un determinado manuscrito se adecuan a unas coordenadas espacio-temporales concretas; además, debe determinar si el tipo de soporte del documento es el adecuado al contenido del texto, incluso si el tipo de datación que aparece en el original se adecua al periodo que se está estudiando, entre otros muchos factores.

(iii) Su menor vinculación a una norma estética –lo que no implica, por supuesto, que estos textos no sigan ciertos moldes y rutinas, como veremos más adelante–, circunstancia que permite que aparezcan en ellos fenómenos que aún no pertenecen al registro culto/literario, pero que existen ya en la lengua.

(iv) La posibilidad de cotejar muchos ejemplos en diversas geografías y sincronías. Si en la literatura se cuenta (no en todas las lenguas, como ya hemos visto) con varios cientos de nombres para realizar tal descripción, en el caso de los archivos el número de autores se multiplica de forma exponencial, puesto que son muchas las manos que dejan testimonio escrito. Es indudable, por tanto, que este hecho permite construir un panorama más completo y aproximado a lo que sería la descripción de una lengua en un momento pasado. Esto es, pese a tener una formación parecida no producirá el mismo tipo de textos un notario que pertenezca a la Cancillería Real que un notario que ejerza su oficio en una población pequeña, del mismo modo que el tenor, la finalidad, el formato y el tipo de receptor serán totalmente distintos.

Cabe concluir, a partir de todo lo mencionado, que es indispensable contar con la mayor cantidad posible de testimonios para ser capaces de dar cuenta de la variación inherente en cualquier estado de lengua; y esto solo es posible, como ya hemos apuntado, con un corpus de trabajo que no utilice únicamente textos literarios, sino que contenga ejemplos procedentes de alguno de los numerosos archivos históricos existentes para el estudio de sus lenguas.

2.1. El tipo de archivo

Los archivos son la memoria de las personas y de las instituciones y existen desde el mismo momento en el que el hombre comenzó a fijar por escrito sus relaciones como un ser social, para recordar y perpetuar los hechos acontecidos y sus actividades públicas y privadas. Aragón nos ofrece un extraordinario patrimonio documental a la espera de que los estudiosos de diferentes ámbitos del saber lo analicen y estudien para darlos a conocer.

La Corona de Aragón en general, y el reino homónimo en particular, es un espacio privilegiado en este sentido. Debemos remontarnos al surgimiento del notariado profesional durante los siglos XIII-XIV en Aragón para entender cómo este tipo de escrituras prosperan y se generalizan. Si comparamos este territorio con otros europeos, puede comprobarse que es quizás el que mayor número de series de inventarios notariales y otros documentos conserva en sus archivos, ya que, a diferencia de lo que sucede, por ejemplo, en Castilla, la estricta legislación foral aragonesa aseguró la transmisión de la documentación notarial de generación en generación. Su preservación hasta nuestros días está estrechamente ligada a la conservación de los protocolos o registros que elaboraba cada notario con las escrituras que expedía, con la finalidad de garantizar su autenticidad.

El tipo de archivo condicionará la búsqueda que el investigador quiera desarrollar. Uno de los más interesantes relativo a nuestro territorio es el Archivo de la Corona de Aragón, y lo es porque excede el ámbito de una Comunidad Autónoma, puesto que se trata del depósito de la historia de los territorios de la Corona de Aragón. Fue creado como decisión soberana de Jaime II de Aragón en 1318 con la consideración de Archivo Real, como propiedad estricta del monarca durante siglos; de hecho, hasta 1770 estuvo alojado en el Palacio Real de Barcelona. Junto con las escrituras referentes al Patrimonio Real se custodiaron allí los documentos de gobierno y justicia, y entre ellos las series de registros de la Cancillería. Pronto creció la complejidad de la oficina: en 1346 el Rey nombraba su primer archivero, con este preciso nombre, y en 1384 le daba normas prácticas sobre la incumbencia de su cargo. Empezar un estudio filológico en un archivo de estas características supone un ejercicio previo de indagación sobre la documentación y de localización de manuscritos concretos en fondos y series determinadas, puesto que utilizar en exclusiva criterios como el de la lengua en la que se redactó el documento o el territorio en el que fue producido, podría resultar un fracaso absoluto.

Por otro lado, las instituciones eclesiásticas han recogido y conservado su propia documentación eficazmente y se han convertido en una referencia obligada, no solo para la historia eclesiástica, sino también para la civil. Esta cuestión tiene diversas explicaciones; se debe destacar la continuidad y la buena organización de estas instituciones –papado, obispado, monasterios, etc.– cuyos miembros tenían un nivel cultural superior, hecho que favoreció tanto la producción como la conservación de sus escritos. Los archivos, herederos del mundo romano, tuvieron un lugar muy importante en la estructura estatal, que más tarde heredó la Iglesia católica, basada en preceptos sagrados conservados y transmitidos mediante la palabra escrita y con una predisposición clara hacia la conservación y custodia de los documentos (Pagarolas 1999: 158).

Entre los archivos catedralicios que custodian documentación relativa a poblaciones de la Franja de Aragón, podemos mencionar el capitular de Lleida (ACL) y el de Tortosa (ACT). Un archivo catedralicio pertenece a la iglesia catedral de una diócesis y custodia la documentación producida y recibida por la comunidad en el ejercicio de sus competencias y en relación con sus derechos y deberes. Los textos que normalmente pueden consultarse son los relativos a su propio gobierno, la gestión del patrimonio, el culto a la catedral, etc.

De esta manera, para el conocimiento de la historia y la lengua en la Ribagorza, la documentación del archivo capitular de Lleida es fundamental, puesto que en él está accesible la documentación generada por la antigua catedral de Roda de Isábena (Huesca), cuyo fondo, con el mismo nombre, se trasladó a Lérida en 1864. Los manuscritos de la catedral rotense son, mayoritariamente, transacciones económicas entre el monasterio y los campesinos de poblaciones vecinas que arrendaban sus posesiones a cambio de una parte de la cosecha o a cambio de una compensación económica. Se trata de textos que contienen establecimientos enfitéuticos, ventas, albaranes, censales, etc.; además, la catedral asumió la documentación de otras parroquias, lo cual hace que tengamos a nuestra disposición un fondo interesantísimo y muy variado. La mayor parte de la documentación a partir del siglo XIV se encuentra redactada en catalán y datada en las poblaciones ribagorzanas cuyas tierras estaban bajo el dominio del obispo de Roda, primero, y del capítulo de Roda y su prior, más tarde (Moret y Tomás 2014). Es menor la documentación escrita en aragonés, pero hay verdaderas joyas para los lingüistas que dan buena cuenta de la realidad lingüística actual de la zona, como manuscritos judiciales donde se muestra mezcla de lenguas (catalán y aragonés).

También encontramos documentación relativa a las poblaciones turolenses de Calaceite, Arenys y Lledó en el Archivo de la catedral de Tortosa (ACT), en el que puede localizarse variedad de cuestiones tratadas y gran número de documentos medievales, la mayoría en pergamino, que hacen referencia, igual que los anteriores, a actividades comerciales, protestas y acuerdos, entre otros.

En otros monasterios, como el de Poblet, en concreto en l'Arxiu de la Casa Ducal de Medinaceli a Catalunya (ADMC), se encuentra la mayor parte de la documentación pública y privada de la Casa de Aitona de la familia Montcada, entre cuyos dominios estaban las poblaciones de Fraga y Mequinenza, donde hay documentación muy interesante sobre las relaciones de posesión de esta poderosa familia noble con la comunidad de sarracenos de dichas villas de la Franja de Aragón (véase Monjo 2004).

Son de similar riqueza los archivos históricos, cuya función es la de mantener la unión de los manuscritos. Este principio se interpreta mediante la unidad del proceso documental y en el hecho de que nunca deben considerarse al margen de la institución o el órgano que ha generado sus fondos. Es decir, un documento es susceptible de convertirse en histórico, de formar parte de su patrimonio documental, desde el momento de su producción. El conjunto orgánico de manuscritos será, por tanto, histórico cuando haya perdido su vigencia administrativa, no haya sido objeto de dictamen de eliminación y haya adquirido un valor permanente, del tipo que sea, preferentemente secundario, cultural o histórico.

En Aragón contamos con diversos archivos de esta índole y se puede consultar documentación redactada en catalán en el Archivo Histórico de Fraga (Huesca); fuera de

esta comunidad autónoma, en el Arxiu Comarcal del Baix Ebre (ACBE) en Tortosa. Del primero, aparte de la colección de pergaminos que abarca los siglos XIV-XVI redactados en distintas lenguas, son especialmente relevantes las actas (desde hace unos años en el Archivo Municipal de la población), ya que en ellas se observa la sustitución abrupta y el abandono de la lengua catalana en las instituciones a mediados del siglo XVI (Moret y Sorolla 2018). En el segundo se conservan diversos fondos con documentación relativa a toda la parte de la Franja de Aragón que de una manera u otra estuvo relacionada con el comercio fluvial a lo largo del río Ebro (Moret 2010).

Los archivos municipales son posiblemente los más interesantes, por el tipo de fuentes que atesoran, cuya creación es coetánea al surgimiento de los centros urbanos medievales. Aquellos que han conseguido conservar los documentos que allí se generaron albergan documentación desde el siglo XIII, forman una unidad diplomática que constituye una parte importante del patrimonio histórico y una referencia inagotable de datos para explicar su pasado. Los manuscritos generados por los antiguos consejos han sido y son uno de los núcleos de información básica para el conocimiento de los movimientos sociales en todas sus dimensiones, y excepcionales desde un punto de vista lingüístico, puesto que son en los que de manera más palpable podemos encontrar huellas diatópicas. Algunos de los que mayor producción manuscrita medieval han conservado son los de las poblaciones turolenses de la Comarca del Matarraña (Teruel), como los de La Fresneda, Calaceite, Fuentespalda o Mazaleón (véase Giralt y Moret 2018), pero también los de otras zonas más al norte de la Franja, como los de Mequinenza, Peralta de la Sal, Albelda o Roda de Isábena (Giralt 2010, 2012; Moret 2010), cuya documentación está mayoritariamente escrita en catalán.

Por último, es necesario mencionar los archivos familiares, creados individualmente por personas, cuyos documentos pasan a las generaciones sucesivas. Pueden incorporar también fuentes ajenas como consecuencia de diferentes transacciones derivadas de negocios familiares u otros intereses. Pero, lo más interesante de estos fondos es que, como mayoritariamente se encuentran en manos privadas y carecen de organización y de instrumentos descriptivos, permanecen inéditos. Estos archivos son quizás los más desconocidos y los que entrañan alguna dificultad a la hora de consultar sus fondos, puesto que a menudo no son accesibles y con frecuencia carecen de organización y catalogación de los manuscritos que los conforman.

Un claro ejemplo de ello es el archivo del Barón de Valdeolivos ubicado en la que hoy es la casa museo de la población de Fonz (Huesca). El sexto barón, Francisco Otal y Balonga, guardó ordenadamente sus escritos y los de sus antepasados; obsesionado por aumentar el archivo y la biblioteca familiar, reunió documentación medieval relativa a municipios de la zona (algunos hoy despoblados) redactados en diversas lenguas, entre ellas el catalán y el aragonés (Moret 2010).

En definitiva, es más que evidente el papel esencial que desempeña el archivo como organizador y gestor de esa memoria y, para comodidad del investigador, la visión del archivo tradicional y la forma de consultar su contenido han evolucionado en las últimas décadas. Las nuevas tecnologías nos ofrecen una oportunidad histórica para romper con la imagen de espacio cerrado y lejano que desde siempre nos han transmitido. Todo ello incluso está modificando los hábitos de investigación. Uno de los servicios más valorados, actualmente, es el acceso a los instrumentos de descripción con la posibilidad de consultar desde cualquier punto y a cualquier hora la información que custodia el

archivo. Desde hace unos años los estudiosos lo tienen más fácil gracias a los llamados “archivos en línea”, que hacen posible el acceso remoto a los fondos documentales, de manera que *stricto sensu* se puede hablar del proceso de “desmaterialización” del archivo: la conversión de los archivos tradicionales en archivos virtuales, abiertos las 24 horas los 365 días del año. Gracias a esto, hoy es posible localizar con un solo golpe de tecla, y a través de una sola operación de búsqueda, todos los documentos existentes en cualquier archivo sobre un determinado tema de investigación.

En este punto podemos mencionar el portal *PARES (Portal de Archivos Españoles)* del Ministerio de Cultura y Deporte del Gobierno de España, destinado a la difusión en Internet del patrimonio histórico documental español conservado en su red de centros. En el caso de Aragón encontramos el portal web *DARA (Documentos y Archivos de Aragón)*, a través del cual podemos acceder a 72 archivos de nuestra región, entre ellos del Gobierno de Aragón, de históricos provinciales, de diputaciones provinciales, de ayuntamientos, de comarcas, religiosos, de asociaciones y de fundaciones o universidades. También permite realizar búsquedas temáticas, como la de *DOMA (Documentos Medievales Aragoneses. De Ramiro I a Fernando II)* que incluye, a su vez, materiales de 45 archivos. En su conjunto, 22.000 documentos anteriores a 1516, de los que cerca de 9.000 se pueden consultar íntegros en red con más de 100.000 páginas digitalizadas. El último avance, iniciado en 2018 por iniciativa de la Dirección General de Política Lingüística del Gobierno de Aragón e implementado en 2020, es la posibilidad de poder localizar manuscritos según la lengua de redacción de los documentos: aragonés y catalán. Asimismo, a finales de 2021 estará activo el portal web *BIVIRA (Biblioteca Virtual del Aragónés)*, donde se reunirán todos aquellos documentos dispersos en lengua aragonesa desde la Edad Media hasta los años setenta.

Lo que es indudable es que el “archivo”, como lugar de conservación de testimonios escritos, es una necesidad de toda organización administrativa, sea cual sea su naturaleza –religiosa, económica, etc.–, sobre todo si se trata de instituciones que han desarrollado un sistema burocrático y, en este caso, deviene esencial tanto por las exigencias puramente administrativas como por la necesidad de conservar la memoria histórica de la institución.

2.2. La tipología documental

Si con anterioridad se ha planteado como problema el empleo de los textos literarios como base de estudios filológicos por su alejamiento de la expresión cotidiana y por su eminente finalidad estética, cuestiones semejantes podrían esgrimirse para la documentación notarial, habida cuenta de la distancia existente entre el tipo de lengua que presentan estos textos y aquella que se utiliza de forma oral. Sin embargo, defenderemos que esta afirmación es sólo parcialmente cierta. En efecto, no cabe ninguna duda de que la lengua que aparece en testamentos, ventas y ordenanzas sufre, al igual que en los textos literarios, un importante proceso de elaboración formal que los sitúa en un polo bastante alejado de la esfera comunicativa oral; ahora bien, mientras que en el caso de la literatura ese alejamiento se advierte en todo el texto y se manifiesta en toda su extensión, en los documentos notariales, no solo se concentra en unas partes muy determinadas, sino que, además, es posible identificarlo y aislarlo con cierta fiabilidad, dada la estructura rígida, rutinaria y repetitiva que ofrecen estas tipologías textuales a lo largo del tiempo.

Para comprender cómo se podía mantener la estabilidad de esta lengua, debemos tener presente que los notarios recibían una formación específica común y debían superar una prueba idéntica para ejercer su oficio. A ello se sumaba la amplia difusión de formularios notariales donde aparecían las expresiones estereotipadas de los documentos. El problema, pues, no estriba en este caso en el texto en sí, sino en la necesidad de desarrollar un método adecuado que permita emplearlo como base en los estudios históricos.

Es necesario señalar, con todo, que no siempre ha existido unanimidad entre los estudiosos sobre el interés que encierran estos documentos en las investigaciones lingüísticas de perspectiva histórica. Mientras que algunos miraron con reticencia los datos que esta documentación ofrecía –y consideraron que carecía de importancia para el estudio de la historia de la lengua por su aparente carácter poco realista–, otros muchos los aceptan y defienden por razones muy diversas, e incluso ponderan su cercanía a la realidad lingüística, al indicar que han proporcionado el material más seguro para la investigación de la dialectología diacrónica, pues la lengua en que suelen estar redactados está más cerca siempre de las formas vulgares que la lengua de los textos literarios. Son opiniones como esta, por tanto, las que justifican y explican que, en el momento actual, un gran porcentaje de los estudios de historia de la lengua utilicen como corpus de trabajo este tipo de textos. En este punto, es necesario señalar que no se trata de una tendencia estrictamente actual.

Volviendo a cuestiones de tipo metodológico, podría afirmarse que, en general, cualquier texto escrito resulta interesante para el estudio histórico de la lengua siempre que se aplique en el análisis un método adecuado a sus características; pues bien, esta idea es especialmente relevante en el caso de los documentos notariales, y de hecho se puede decir que esta documentación adquiere una importancia esencial para estos estudios, hasta el punto de ofrecer datos que muy difícilmente se encuentran en las otras tipologías ya mencionadas.

Entendemos por *documento notarial* un texto de carácter legal que llega a ser vinculante al establecer una relación entre dos partes o más de diversa índole: compraventa, donación, herencia, etc. Representa la tradición discursiva propia de la distancia comunicativa; esto es, se caracteriza, en principio, por una marcada planificación textual, por la que cada parte del texto dispone de unas fórmulas fijas, primero en latín y desde finales del siglo XIII en romance, modernizadas con el paso del tiempo (García-Valle 2004: 629), factor que no le impide al redactor introducir innovaciones lingüísticas de carácter popular.

Bajo el paraguas de “documentación notarial” encontramos gran variedad de registros (narrativo, testamentario, contractual, etc.) y tipos de soporte en los que fueron escritos, circunstancia que también puede aportarnos información de distinta naturaleza. Sin entrar en cuestiones puramente lingüísticas, *grosso modo*, podemos afirmar, por ejemplo, que los pergaminos contienen una *scripta* más conservadora, por ajustarse más a la del catalán general de la época, mientras que los protocolos (libros en papel) presentan más soluciones dialectales y nos permiten aproximarnos con mayor detalle a la realidad material de las gentes de la zona, aunque no siempre tiene que ser así, como veremos; por ello, dependiendo del tipo de estudio que queramos realizar, podremos acudir a una u otra fuente.

Son de especial interés para determinar este tipo de fenómenos las cartas, requerimientos y protestaciones, que también requerían de la destreza de un notario para transcribir el asunto que se quería dejar por escrito para que se diera traslado a la autoridad pertinente. No es extraño encontrar expresiones coloquiales en un texto completamente formal, como en los siguientes ejemplos (véanse Figura 3 y Figura 4), en los que puede leerse la expresión «a ull» ('a ojo, con los propios ojos', DCVB, s.v.: *ull*):

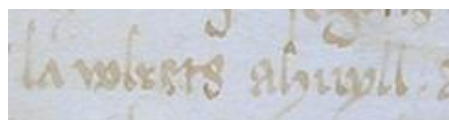
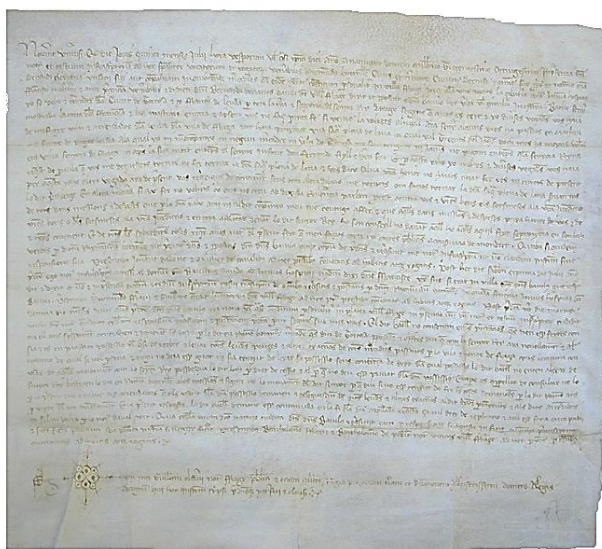


Figura 3. 1380, 5 de juliol. Fraga.ACBE. Requestes i protests I 28. Perg. 587. Not. Guillem Oliver

«[...]Sènyer, yo só vehí e mercader de la ciutat de Tortosa e só ffranch de leuda per tota la terra e senyoria del senyor rey d'Aragó, segons que a vós és cert e yo diverses vegades vos haia mostrada la carta de la ffranquesa e la us mostraré encara e a present vos ne faç promta fe, si veure la volrets **a huyl** ara [...].»

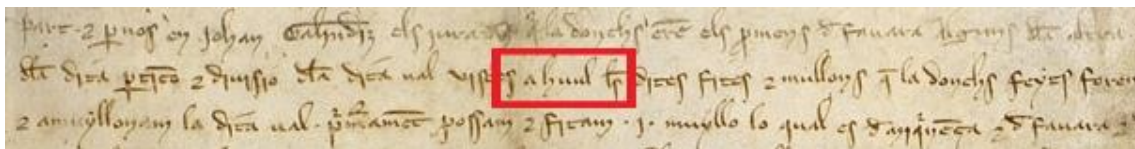


Figura 4. 1331, 23 de novembre. Mequinenza. AMM. Perg. 50. Not. Domingo Jordà: «[...] segons que és contengut en cartes públiques feytes per en Domingo de Plaça, notari de Miquinença, ara a mayor e a myllor declaració de la dita partició e divisió de la dita val, vistes **a huyl** les dites fites e mullons que ladonchs feyts [...].»

En conclusió, podem senyalar que, aunque el documento notarial no deja de ser un texto escrito más, como lo es el literario, su finalidad social es completamente distinta y esa circunstancia ha favorecido que afloren soluciones lingüísticas dialectales que reflejan la lengua viva de aquellas sincronías pasadas y que justifican la configuración actual de las variedades diatópicas de la Franja de Aragón.

2.3. La escritura de los profesionales medievales

Dejando de lado los registros de la cancillería real, no es hasta finales del siglo XIII cuando se atestiguan las primeras muestras de textos completamente escritos en catalán en estos territorios. Es en esta centuria –en la que la realidad lingüística de la Corona de Aragón se manifiesta en la producción escrita y en la que se pone en evidencia la influencia del bilingüismo institucionalizado de la cancillería real– cuando se generalizó el uso en la escritura del catalán y del aragonés en estos territorios, aparte de seguir

usando el latín con una función predominantemente litúrgica o como lengua franca (Tomás 2020).

No resulta extraño, por tanto, localizar manuscritos en catalán, sobre todo, desde principios del XIV, y este hecho resulta más generalizado a partir de la segunda mitad de siglo, en el cual los escribanos de la chancillería real, por un mandato del rey, estarán obligados a dominar las tres lenguas de la corona (catalán, aragonés y latín), hecho que no debía estar sujeto a que el uso de diversas lenguas en la documentación consistiera en el conocimiento espontáneo de las mismas por parte de los notarios, particularmente, en relación al latín, lengua aprendida para ser escrita o para la lectura en voz alta (Moret 2020: 129).

Sin embargo, no sería justo asegurar que en esta época encontraremos el reflejo de la lengua oral, puesto que la realidad es mucho más compleja y se debe ir con cautela a la hora de poner en relación escritura y oralidad. El romance en documentación es una convención, una lengua adquirida para ser escrita y no refleja –o por lo menos no tiene la intención de hacerlo– la lengua oral, excepto en contadas ocasiones.

La lengua usada por los escribanos es la llamada técnica o de especialidad, cuya función es facilitar la tarea del notario y llevar a cabo con eficacia los contenidos de los que tratan los documentos. Este registro, desde una perspectiva formal, en el caso de la lengua jurídica, se caracteriza por unos usuarios concretos (legisladores, autoridades administrativas, tribunales y miembros de las profesiones jurídicas) y unas instituciones de comunicación determinadas: relaciones entre el poder legislativo y judicial y los ciudadanos, así como las existentes entre particulares con trascendencia jurídica.

Desde la alta Edad Media el arte de componer una escritura y el estilo era materia de instrucción regular que determinaba el desarrollo de una retórica documental. Los notarios se servían de esquemas fijos en la elaboración de sus escritos, que habían de estar redactados con precisión y el contenido de estas fórmulas se repetía en cada texto, lo cual facilitaba la redacción a este colectivo de especialistas.

No se debe olvidar, sin embargo, que la trascendencia social del fenómeno foral en Aragón durante toda la Edad Media tuvo como consecuencia que la lengua jurídica no se constituyera como un hecho hermético, sino que se desarrollara una comunicación en ambas direcciones, de un elitismo especializado a una generalización idiomática. No es difícil, por tanto, encontrar en los registros elecciones gráficas entre algunos notarios, separaciones de sílabas, determinadas formas y léxico dialectal, etc., e incluso correcciones continuadas que pueden aportarnos datos en cuanto al cambio lingüístico de un determinado elemento, como el que aparece en la Figura 5, en el que se refleja la sustitución en la evolución semántica de los verbos *ir* y *venir* que, como verbos de movimiento, han experimentado a lo largo del tiempo tanto en catalán como en aragonés (Moret et al. 2017; Ibarretxe-Antuñano et al. 2017; Moret e Ibarretxe-Antuñano 2019):

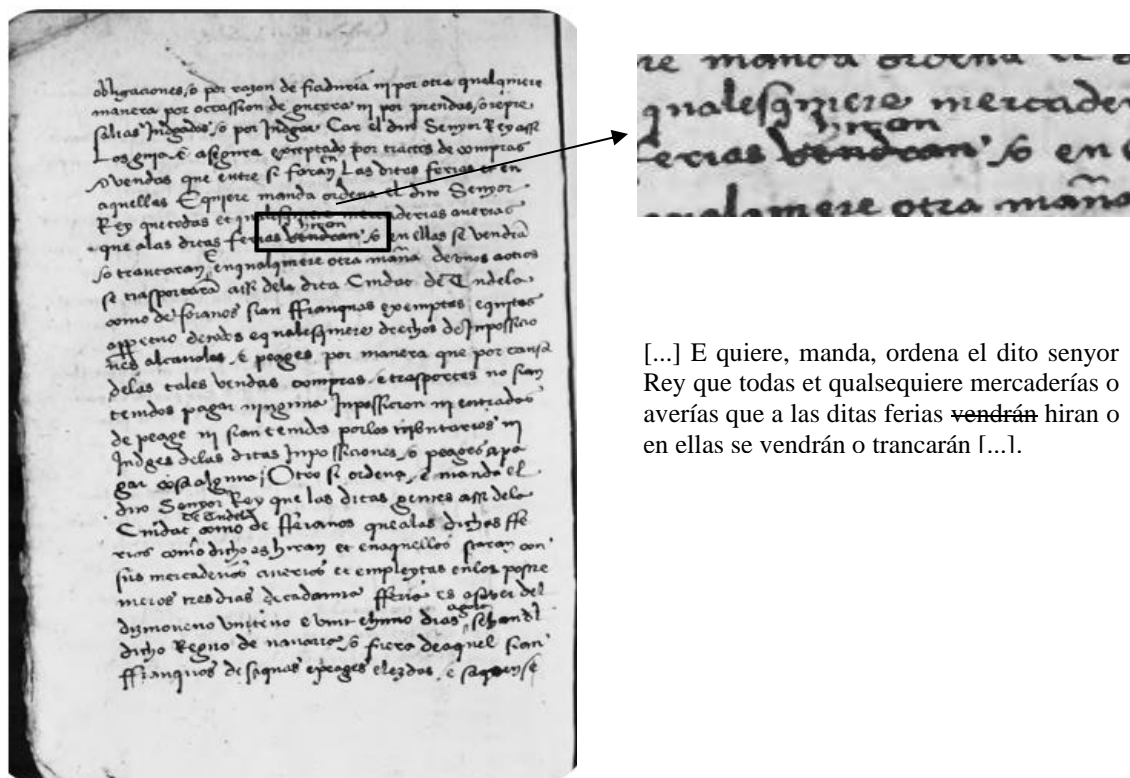


Figura 5. Sustitución del verbo *venir* por *ir* en documento en aragonés, seguramente por influencia de la lengua castellana dada la datación de este documento (1450-1480, *Libro de cridas o pregones*, Zaragoza)

[...] E quiere, manda, ordena el dito senyor Rey que todas et qualsequiere mercaderías o averías que a las ditas ferias **vendrán** hiran o en ellas se vendrán o trancarán [...].

3. Cuestiones metodológicas

Si a lo largo de este trabajo hemos justificado la importancia de este tipo de documentación para el estudio de la lengua, en adelante nos detendremos en las posibilidades que estos corpus presentan para su estudio. Sánchez-Prieto (2012: 9) señala que el historiador de la lengua puede servirse de los corpus textuales, primero, para atestiguar, ilustrar y ejemplificar aquello que ya conoce; segundo, para cuestionar lo que ya se sabe, contrastándolo con los datos empíricos, a través de análisis cuantitativos y cualitativos en los que se tenga en cuenta la cronología de dichos textos, su localización, los aspectos relacionados con el tipo de registro y los aspectos diafásicos y sociolingüísticos que podamos extraer; y, por último, para establecer su tipología textual.

En las siguientes páginas nos centraremos en dos de las fases que pueden llevarse a cabo en un corpus de documentos; por un lado, la edición de textos y las dificultades que entraña, y por otro lado, el análisis de los documentos desde diferentes perspectivas.

3.1 La edición de textos

Un corpus será válido y aprovechable si el historiador de la lengua tiene un control de su edición, o conoce todos los procesos que se han llevado a cabo hasta su resultado final. Todo proceso de edición implica un factor de interpretación y de pérdida de datos (una fase de cualquier proceso ecdótico), puesto que cuando nos enfrentamos a presentar al lector un texto de estas características, procedemos a la eliminación

voluntaria de una serie de elementos que pueden considerarse de escaso valor para ciertos tipos de análisis, como los lingüísticos: abreviaturas, firmas de los notarios, separación de columnas, cambios de línea, etc.; y se toman decisiones, como acentuar el texto, separar o unir palabras, entre otras.

Un historiador de la lengua puede trabajar a partir de dos tipos de corpus: por un lado, los propios, es decir, en los que el investigador cumple todo el proceso, desde su localización y transcripción hasta su edición; y, por otro, los ajenos, en los que se sirve de textos ya editados y puestos a disposición a través de ediciones previas, publicadas o disponibles en línea. Ambas, aunque igualmente legítimas, presentan una perspectiva muy diferente del investigador ante el texto, puesto que en el primero de los casos se tiene un control absoluto sobre las intervenciones acometidas, y en el segundo no tiene por qué ser así. En la actualidad, sin embargo, existen criterios que están bastante unificados en las distintas ediciones de textos; de hecho, la mayor parte de las decisiones que atañen a cuestiones de regularización gráfica se repiten en la mayoría de las ediciones críticas que se hacen actualmente.

Cabe señalar que en el campo de la edición de textos se ha avanzado mucho; en el caso del español existen proyectos panhispánicos, como CHARTA, que engloba la edición y el análisis lingüístico de textos archivísticos en español de los siglos XII al XIX, de Europa, América y Asia, y que no solo acoge piezas oficiales de la cancillería, la administración civil, la de justicia, la Inquisición o contratos de compraventa, sino también cartas particulares y aún billetes y notas sueltas. Por ahora, por tanto, se ofrece una selección amplia de archivos españoles y una breve muestra de documentos americanos, que se ampliará en ediciones sucesivas del corpus. Sus publicaciones, además, presentan una sólida fundamentación filológica con los últimos desarrollos informáticos necesarios para la explotación del corpus. Para facilitar su consulta por diversos usuarios ofrece tres presentaciones de cada texto: facsimilar, paleográfica y crítica, esta última de fácil lectura.

En el caso del catalán las ediciones más consultadas son las de la colección *Els nostres clàssics* (ENC), un proyecto fundado en 1924 por Josep Maria de Casacuberta que, en la actualidad, continúa activo con un extenso catálogo realizado por editores de referencia –aunque también con determinadas críticas–, cuya edición ha ido evolucionando y se ha adaptado a destinatarios y a diversos objetivos, teniendo en cuenta el momento histórico y las circunstancias en las que se ha desarrollado el trabajo editorial (véase Santanach 2012). El proyecto tenía como objetivo facilitar a los lectores catalanes el acceso de textos antiguos en lengua catalana y difundir su conocimiento, además de proporcionar modelos lingüísticos. De hecho, en las primeras ediciones, como la de *Lo somni* de Bernat Metge, editada por J. M. Casacuberta el año 1924, se optó por una regularización ortográfica de la obra a partir de la normativa de Fabra, como el uso de apóstrofes, guiones, acentuación, puntuación, uso de mayúsculas y minúsculas, desarrollo de abreviaturas, etc. Es decir, se presentaba una obra al lector en la que este no encontrara divergencias sustanciales con las normas propuestas por el Institut d'Estudis Catalans. Sin embargo, y a pesar de que estos criterios se mantienen en los primeros tomos de la colección, más adelante se da un giro a la edición de estos textos, en concreto se cambia de rumbo con el *Llibre de les dones* de Jaume Roig, cuya edición la llevó a cabo Francesc Almela Vives. En este caso, se conservan, en general, las grafías del original. En esta obra, por tanto, ya no se regulariza sistemáticamente de acuerdo con la normativa, sino que únicamente se incide en determinados fenómenos como la

unificación de *i/j*, *u/v*, supresión de *h* no etimológica, normalización de todas las grafías que representen la *s* sorda, *s/ss/c/ç*, simplificación de consonantes dobles iniciales, la *l* palatal se representa *ll*, i la *ch* cuando es africada se sustituye por el dígrafo *tx*, etc. La evolución en la edición de los textos en la colección ENC no termina aquí, sino que en las ediciones posteriores se optó por ser más fieles a los originales, con un mayor grado de especialización, mayor extensión en las notas introductorias y en las que aparecen especificadas al final del tomo para no interferir en la lectura, las cuales indican cuestiones sobre referencias culturales, históricas o incluso lingüísticas, así como explicaciones sobre decisiones en la transcripción. Esto deviene, en definitiva, en ediciones menos accesibles a las que se habían ideado en un principio, pero más fieles al texto original y, por consiguiente, más útiles para el investigador.

En nuestras ediciones de manuscritos de la Franja de Aragón escritos en catalán, hemos optado por seguir la nueva orientación que se ha tomado en ENC. No obstante, y teniendo en cuenta que nuestro cometido no es únicamente la transcripción y presentación de los textos, este paso ha supuesto únicamente una fase de nuestro trabajo, puesto que nuestro objetivo final se centra en realizar caracterizaciones lingüísticas de una geografía concreta, en las que abarcamos diversas perspectivas, como la descripción gráfica, morfológica, sintáctica o léxica de los manuscritos. Por esta razón, hemos apelado a la cautela a la hora de tomar decisiones referentes a la regularización en nuestras ediciones, incluso algunas de ellas han tenido que ser reformuladas en los sucesivos trabajos.

I. CRITERIS DE TRANSCRIPCIÓ I EDICIÓ

1. Els criteris que hem utilitzat són els habituals en aquest tipus d'investigacions, basats sobretot en els aplicats en la col·lecció *Elis Nostres Clàssics* de l'Editorial Barcino (veg. Martínez, 1999: 93-98):

a) Hem regularitzat les grafies paleogràfiques *i/j*, *u/v*, i hem simplificat sistemàticament les geminades *ff* i *rr* en posició inicial de paraula. A més, hem escrit en majúscula tots els noms de font, els que hem considerat ja veritables cognoms, així com els topònims majors i menors.

b) Hem usat els signes de puntuació que hem considerat adients, tot i que som conscients que hem pogut incórrer en arbitrarietats, perquè les pauses prosodiques no sempre es corresponen amb les pauses gràfiques.

c) Hem desenvolupat les abreviatures i ho hem indicat posant en cursiva la part elidida en l'original.

d) Els fragments en llatí o en castellà apareixen escrits igualment amb lletra cursiva. Hem d'assenyalar, referent a això, que la transcripció s'ha dut a terme en tots els manuscrits, excepte en el cas del document núm. 14, perquè en català només apareix el memorial de la roba en un esborranç, redactat, a més, per una mà diferent a la del notari (suposadament la que redacta tot el text en castellà).

e) Hem accentuat les paraules d'acord amb les actuals regles d'accentuació gràfica del català. En alguns vocables que es poden prestar a confusió, hem usat l'accent diacrític: *à* = 'ha', *é* = 'hé', per exemple.

f) Hem separat, sense indicar-ho, els mots que en els manuscrits anaven junts i hem respectat separacions, com en el cas de *cosit geronú*. En canvi, hem unit els elements de paraules compostes, del tipus *onsevalle*, quan s'han escrit separats. A més, hem respectat les fusions *della*, *delles*.

g) Hem emprat el guionet i l'apòstrof segons les normes ortogràfiques del català actual, tant en les formes de l'article com en les dels pronoms febles al costat del verb. Tanmateix, hem respectat les contraccions *al* i *del* quan apareixen davant un mot que comença per vocal. D'altra banda, quan l'elisió vocàlica no s'indica mitjançant apòstrof segons la norma ortogràfica, hem col·locat un punt volat (*que l*).

2. Per dur a terme l'edició dels documents, hem pres algunes decisions amb les quals volem donar una visió dels manuscrits tan fidel a l'original com sigui possible:

a) En la presentació dels textos seguim l'ordre original dels protocols, a pesar que no s'hi hagi seguit sempre un ordre cronològic estricta.

EDICIÓ DEL PROTOCOL D'EN PERE ORIOLA (1308-1429)

I. CRITERIS D'EDICIÓ

En la transcripció hem respectat escrupolosament la lletra en la seva part més substancial i els trets gràfics no rellevants els hem adaptat al lector actual, seguint l'exemple de *Elis Nostres Clàssics*, *Scriptes dialectals* i *Col·lecció Scripta*.³⁰ Donem, doncs, una transcripció interpretativa, però conservem la configuració essencial de la llengua original. Resolem les lletres que estan representades per una abreviació, però no es posen en cursiva. Regularitzem la separació de mots i, per facilitar la lectura del text, adaptem la puntuació, en la mesura que és possible, a les regles modernes, així com les majúscules i minúscules als criteris de l'ortografia actual. Mantenim els hàbits gràfics del manuscrit, encara que hi hem intervingut en algunes ocasions:

- En relació amb la *u* i la *v* gràfiques del manuscrit, destriem si es tracta de vocal o de consonant i donem la lectura *u/v* i *ij* segons el cas.

- Simplifiquem la doble *f* inicial, però mantenim les duplicacions de consonants emmig de mot (per exemple, *coffiz*, *agroniz*, *luffiz*).

- Usem apòstrofs, accents, dièresis i guionets d'acord amb la normativa actual, amb la plena consciència d'aplicar a la llengua antiga un sistema d'accentuació basat en obertures vocàliques del català oriental modern. No obstant això, col·loquem accent gràfic o dièresis damunt la *y*, quan aquesta representa un so vocàlic i no constitueix diftong (veg. *parayz*, *vayna*, *benoyz*); també apliquem l'accent diacrític a mots mosaicol·lòbics que es poden confondre amb altres (*à* 'ha', *è* 'allà', *éi* 'all', *èis* 'ells', *néi*), encara que no estiguin regulats per la norma ortogràfica vigent.

- Amb el punt volat «*l*» indiquem la separació de tots aquells casos d'aglutinació que la normativa actual no accepta; amb l'accent circumflex marquem que una *e* equival a *en* («*e*» = *en e*).

Finalment, assenyalant entre barres «*l*» els afegits del notari, introduïm els

Las ediciones que hemos llevado a cabo hasta el momento (véanse Figuras 6, 7 y 8) han evolucionado en cuanto a la toma de decisiones, como las de no mantener fragmentos rallados o tachados que hoy sí dejamos en la transcripción, puesto que pueden ofrecernos información sobre las vacilaciones gráficas o léxicas del amanuense. Asimismo, introducimos entre corchetes «[]» aquellos fragmentos de difícil lectura o borrados, que nos permiten señalar reconstrucciones por parte del editor; entre antilambda «<>» señalamos errores de repetición; y con las barras encontradas «\ /» incluimos los añadidos entre líneas o al margen.

referim a un text que reproduïx o reconstrueix, amb màxima fidelitat, la voluntat de l'autor a l'hora de publicar la seva obra.

Enfrontar-se a l'edició de textos catalans medievals és haver-se de referir a una pràctica prou assentada i als grans noms de la filologia catalana del segle XX, però també a una pràctica que evoluciona constantment i on la diversitat dels textos, les circumstàncies de la seva transmissió o les necessitats dels lectors han condicionat, freqüentment, el resultat d'aquesta disciplina.

La tradició de textos romànics ha optat pel respecte cap a la grafia antiga i en el cas de català hi ha una sèrie de col·leccions que han seguit les mateixes línies d'actuació en les seves edicions, com ara les normes establertes per Els Nostres Clàssics (ENC) o la comissió editora de les Obres de Francesc Eiximenis (OFE) i el Corpus Biblicum Catalanicum (CBCat), que segueixen amb poques modificacions les normes d'ENC.¹⁰

Els Nostres Clàssics va optar per una intervenció reduïda i el respecte màxim per la forma gràfica dels manuscrits, introduint l'accentuació gràfica segons la normativa moderna, la separació de mots segons els criteris de la normativa fabriana, la regularització d'algunes grafies segons l'ús modern o el respecte d'altres si el manuscrit presentava característiques particulars d'un dialecte (cf. Pujol, 2013: 177-247).

Tenint en compte les indicacions que tenen en comú aquestes col·leccions, hem aplicat els següents criteris per tal de dur a terme la nostra edició:

1. Les abreviatures es resolen i s'indiquen en cursiva. Aquesta operació ha presentat, de vegades, problemes quan es pot resoldre amb dues grafies diferents, cosa que ha fet que s'hagi adoptat un criteri coherent amb el comportament normal del manuscrit o amb la norma moderna (si això no contradïu els hàbits del copista). Aquesta situació afecta, bàsicament, al desenvolupament de l'abreviatura de consonant nasal davant de labial (*n* o *m* davant de *b*, *p* o *m*).
2. S'han regularitzat les majúscules i minúscules segons els criteris de l'ortografia moderna.

10 Es poden consultar les normes en els portals que s'noten a la bibliografia.

3. Hem respectat de manera escrupolosa els hàbits gràfics del manuscrit i hi hem intervingut en poques ocasions, com ara:

- La unificació de les dues grafies de la sibilant sorda (*s* i *f*) en *s* o la normalització segons l'ús modern de *ij* o *uv*, emprades indistintament com a vocals o consonants, que s'assenyalaran amb el seu valor fonètic, tot prescindint de la forma en què apareguin en el manuscrit.
- Hem simplificat la doble *f* inicial, però hem mantingut, en canvi, les duplicacions de consonants enmig de mot (per exemple *bb*: *pubbiles* 8, *púbblie* 8, *abbat* 29, *affegir* 28, etc.).
- Presenta alguns problemes l'ús de la geminada (*ll*) per distingir la doble *l* de la *l* palatal. Hem optat, per tant, pel respecte a la grafia que apareix en el text, ja que podria generar grafies no etimològiques (**mal-la*, **mal-ler*) o impossibles (**el-ls*) o la falsa seguretat de la pronúncia de mots com *mal-la*, *apellar* - *apel-lar*.

4. S'ha determinat fer ús d'apòstrofs, accents, dièresis i guionets d'acord amb la normativa actual, amb la plena consciència d'aplicar a la llengua antiga un sistema d'accentuació basat en obertures vocàliques del caràcter oriental modern. De fet, l'accentuació gràfica pot presentar alguns problemes d'ordre gràfic, per la inestabilitat de les grafies medievals que pot provocar formes estranyes com *renin* 10 (per *renim*), o l'exigència de l'ús d'accentos diacrítics en mots com à (= *ha*) 29, é (= *he*) 2.

5. Els números es reproduïxen en xifres romanes o àrabs, segons apareguin en l'original.

6. El signe «*et*» es transcriu per *et* tant en els fragments en llatí com en romànc.

7. En la transcripció se separen les paraules que en l'original es troben unides i, per tant, s'uneixen les lletres o síl·labes d'una mateixa paraula que apareixen separades.

8. Per tal de facilitar la lectura del text, s'adapta la punnació, en la mesura que sigui possible, a les regles modernes.

9. Es respecten les contraccions dels mots.

10. El punt volat «*·*» indica separació dels pronoms i tots aquells casos d'aglutinació o de contracció que la normativa actual no accepta.

Figura 8. Criterios de edición de Giralte y Moret (2018: 30-31)

Los cambios más significativos respecto al manuscrito original son los que atañen a la regularización de mayúsculas y minúsculas, según los criterios de la ortografía moderna, la separación de palabras que en el original se encuentran unidas o la unión de las que aparecen separadas (que en las primeras ediciones manteníamos), la regularización de la puntuación respecto a las contracciones de palabras que el texto presenta, la normalización según el uso moderno de *i/j* o *u/v*, la simplificación de consonantes iniciales como *ff* (aunque mantenemos las que aparecen en interior de palabra). También los cambios que atañen al desarrollo de grafías o de palabras, como en el caso de las abreviaturas (que marcamos en cursiva, puesto que se trata en algunos casos de hábitos de notarios que nos permiten hacer descripciones detalladas sobre usos gráficos), así como la inserción de un acento circunflejo sobre la *e* en aquellos casos en los que equivale a *en*: «*ê·l*» = *en el* (véase Figura 9), un error de lectura que con frecuencia pasa inadvertido en transcripciones realizadas por historiadores o filólogos.

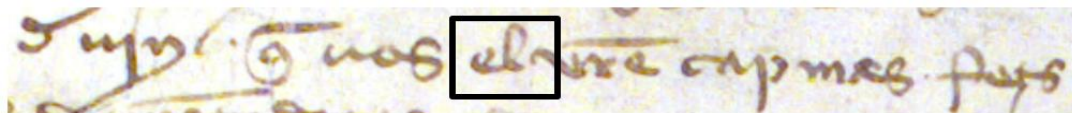


Figura 9. Fragmento del pergamino redactado por el notario Beringuer de Solanella, perteneciente a la antigua catedral de Roda de Isábena, del año 1365, en el que puede observarse la contracción entre la preposición “en” y el artículo “el”: «[...] de vin *que* vós ê-l vostre capmàs feys».

En relación con los primeros trabajos realizados sobre documentación medieval hemos eliminado todos aquellos signos que no aportan información sobre el texto, como la partición de líneas en un pergamino y su numeración para facilitar su lectura, o incluso las elaboradas firmas notariales.

En suma, enfrentarse a una edición de un texto supone una toma de decisiones, es decir, el editor debe ser el encargado de transformar el texto fuente en un nuevo producto que debe preparar para un público, experto o no; y dependiendo de este factor, determinar el grado en el que mantendrá el estado gráfico de la fuente y cómo intervendrá en el texto para su modernización.

3.2 El análisis de los documentos

Los estudios lingüísticos diacrónicos han evolucionado muchísimo, puesto que la lingüística de corpus ha avanzado a pasos agigantados (véase García-Miguel 2022). Por esta razón, cada vez existen más proyectos que introducen documentos transcritos en una base de datos para que cualquier investigador pueda acceder a ellos. Algunos como CICA o CHARTA, que ya hemos mencionado anteriormente, han permitido que trabajos científicos de gran interés hayan visto la luz, puesto que estas bases de datos facilitan la comparación de las distintas variantes de un fenómeno gramatical, de cuestiones gráficas, morfosintácticas, etc., de muchas fuentes y en muchos contextos, de manera que permiten llevar a cabo estudios comparativos en distintas áreas geográficas y diferentes sincronías. El primero de ellos, el *Corpus Informatitzat del Català* (CICA), posibilita el manejo de un número de obras muy elevado y la obtención de una gran cantidad de ejemplos, algo que sería muy difícil si se tuviera que realizar a mano.

En nuestro caso, como único grupo de investigación que se dedica a la documentación aragonesa medieval en lengua catalana, hemos conseguido aglutinar un corpus amplio de manuscritos, datados y redactados en todos los territorios que componen lo que actualmente conocemos como Franja de Aragón, desde el siglo XIV hasta el XVI. En concreto, en las ediciones realizadas hasta el momento hemos cubierto todo el territorio entre estos siglos.

De Ribagorza (Huesca) hemos conseguido aglutinar un corpus, mayoritariamente del fondo del monasterio de Roda de Isábena, con documentos datados en el siglo XIV y XV de Benabarre, Esdolomada, Estadilla, Güel, Graus, Lascuarre, Nocellas, San Esteban del Mall, Serraduy y Roda. De La Litera (Huesca), se han publicado los protocolos notariales del siglo XVI de Albelda. Del Bajo Cinca, hemos recopilado documentación medieval en catalán de los siglos XIV-XVI de Fraga (Huesca) y Mequinenza (Zaragoza). Ya en la zona sur, dentro de la actual comarca del Matarraña (Teruel) hemos estudiado manuscritos en Calaceite, Cretas, Fuentespalda, La Fresneda, Mazaleón, Monroyo,

Ráfales, Valderrobres y de poblaciones cercanas, como Alcañiz (Teruel), Morella (Castellón), Poblá de Benifassà y Tortosa (Tarragona) (véase Figura 10). A continuación, presentamos en los párrafos que siguen algunos ejemplos de aspectos que se pueden estudiar, desde una perspectiva lingüística, en los diferentes niveles de análisis.

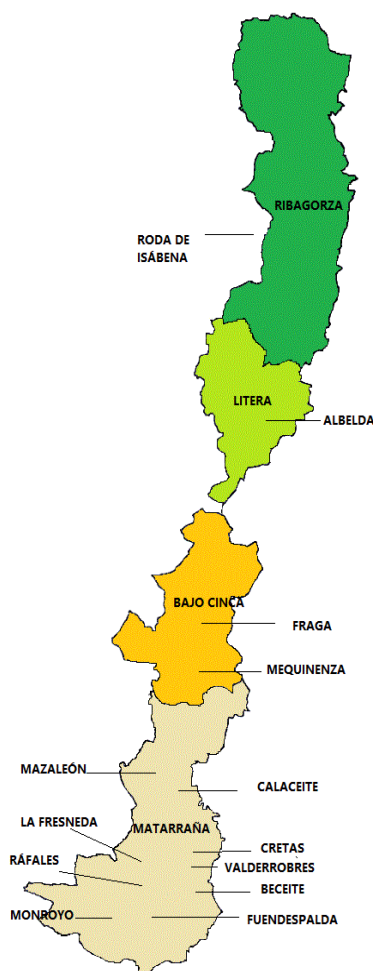


Figura 10. Poblaciones de la Franja de Aragón con documentación estudiada y datada

3.2.2. Grafía y fonética

La compilación de documentos medievales originales de todas las zonas de Aragón redactados en catalán nos ha permitido llevar a cabo descripciones en el nivel gráfico y fonético. Si bien la escritura catalana de estos territorios no difiere especialmente de la que encontramos en otros puntos del dominio lingüístico de toda la Corona de Aragón, sí advertimos en todos los textos estudiados una regularidad gráfica general en todo el XIV y XV con el uso de una gótica cursiva documental aragonesa, cuyos rasgos característicos son el amplio desarrollo de los astiles superiores e inferiores (véanse ejemplos en la Figura 11):



Figura 11. Distintos ejemplos de las letras *b*, *h*, *j*, *m*, *p*, *x*, extraídos de documentación aragonesa en catalán del siglo XIV

Además, en todos estos textos, está presente una perfecta separación entre palabras:

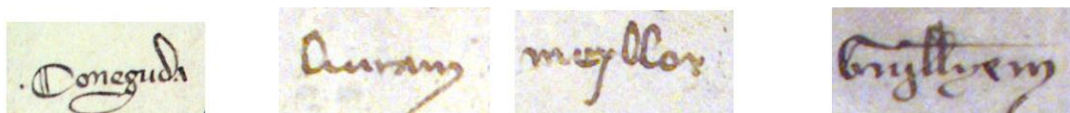


Figura 12. Palabras: *coneguda*, *liuram*, *meyllor* y *Guylliem*, extraídas de pergaminos del siglo XIV procedentes de la antigua catedral de Roda de Isábena, en las que se observa la perfecta separación gráfica que facilita la lectura

Así como, se observan las formas amplias y bajas de las letras, la caída de los palos inferiores de la *s*, *i*, *f*, *p* y *q*:

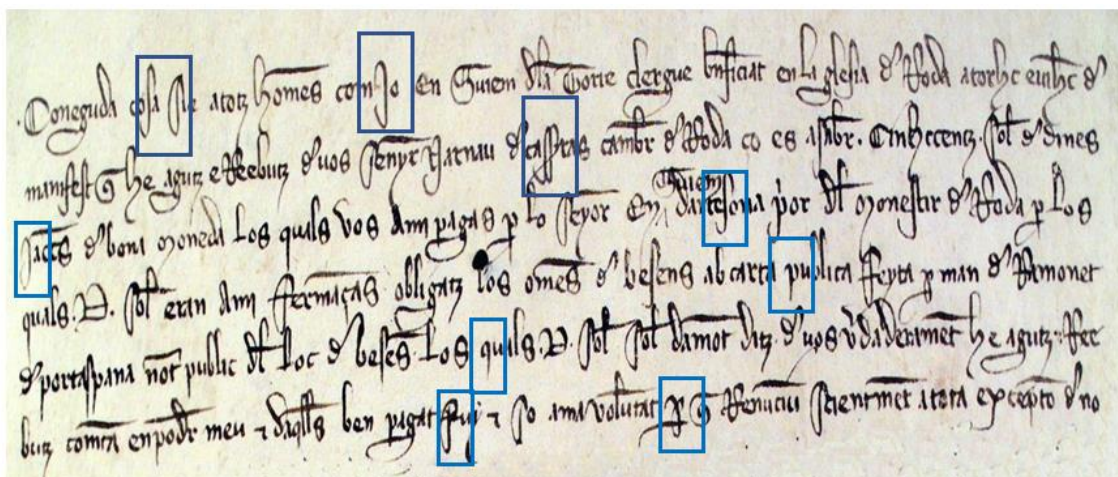


Figura 13. Fragmento inicial del pergamino 175 del Fondo de Roda (Archivo Capitular de Lleida), datado el 2 de septiembre de 1333. En este pergamino pueden observarse los trazos alargados de las letras *s*, *i*, *f*, *p* y *q*, en el que puede leerse: «Coneguda cosa sie a totz hòmens, com jo en Guiem de la Torre, clergue beneficiat en la glésia de Roda, atorhc e vinhc de manifest que he agutz e reebutz de vós senyor n' Arnau de Casserras, cambre de Roda, ço és a saber, cinhc-centz sòlidos de dinés jaqueses de bona moneda, los quals vós a mi pagàs per lo senyor en Guiem d' Artesona, prior del monestir de Roda, per los quals D sòlidos éran a mi fermanças. Obligatz los òmens de Besens, ab carta pública feyta per man de Ramonet de Portaspana, notari públic del loc de Besens, los quals D sòlidos <sòlidos> damont ditz de vós verdaderament he agutz, reebutz, comtan en poder meu et d'aquells ben pagat fuy e só a ma voluntat perquè renunciu scientment a tota excepció de no [...]».

La *g* comienza a esbozar un giro hacia la izquierda que envuelve el trazo inferior y la *r* se mantiene sin adoptar aún la forma típica aragonesa, un irrefutable influjo francés de la letra *batârde* que se da sobre todo durante el reinado de Juan I (1387-1397) y relacionado con las esposas francesas del rey (Riesco 1996: 42):



Figura 14. Distintos tipos de letra *g* y *r*, extraídas de pergaminos de la antigua Catedral de Roda de Isábena del siglo XIV

En general, a pesar de que Aragón presenta una tradición escrituraria bastante uniforme, puede confirmarse, desde una perspectiva escriptológica, que en algunos casos existe otra diferente entre los documentos procedentes del centro monástico de Roda de Isábena y los de zonas situadas más al sur, lo cual no quiere decir necesariamente que se trate de dos variedades dialectales (Goebel 1975).

Si en las líneas precedentes hemos ejemplificado el análisis gráfico de un manuscrito, consideramos que también es interesante demostrar la interacción que existe entre los niveles gráfico y fonético, y qué implicaciones conlleva a la hora de documentar y explicar determinados resultados fonético-fonológicos de las variedades dialectales actuales. En este sentido, traemos a colación dos aspectos a los que Coromines dedicó unas cuantas páginas y que tienen su reflejo en los documentos que hemos estudiado. Nos referimos a la confusión de las consonantes antiguas *ç* (*c+e, i*) / *ss* (*s-*, *cons.+s*, *-s*), y *yl* / *ll*. Las afirmaciones de Coromines (1976: 13-51), para las cuales no poseía documentación de todos los territorios catalanohablantes, podrían verse matizadas y aumentadas con la documentación aragonesa que abarca toda la Franja de Aragón. Coromines (1976: 14) señala que la grafía *ç* (*c* sibilante ante *e, i*) antes del 1200 tenía un sonido africado [ʧ], que hoy ha evolucionado a un *s* sorda en todo el territorio lingüístico, excepto en algunas zonas de frontera lingüística en la Ribagorza (Huesca) y el Bajo Aragón (Teruel), al igual que el valle gascón de Vatmala, donde ha evolucionado a [θ], como en aragonés y castellano.² Plantea Coromines (1976: 16-17) que la confusión entre [ʧ] y [s] se produjo ya en el siglo XIII en la zona costera catalanohablante, desde el Rossellón hasta Valencia. A partir de entonces, la isoglosa de la confusión entre los dos sonidos se fue ampliando hacia el oeste, de manera que en el siglo XIV había llegado ya a tierras leridanas, quedando la distinción entre la africana y la fricativa reducida a lugares más aislados, tanto del Pirineo catalán como del maestrazgo valenciano, y en la actual Franja de Aragón. Es a partir del siglo XV cuando la confusión se va generalizando en esos territorios del área occidental del catalán, salvo en las zonas en las que actualmente hoy todavía se articula [θ], es decir, las que ocupan las variedades del Bajo Aragón y las hablas de transición de la Ribagorza y La Litera (véase Figura 15).

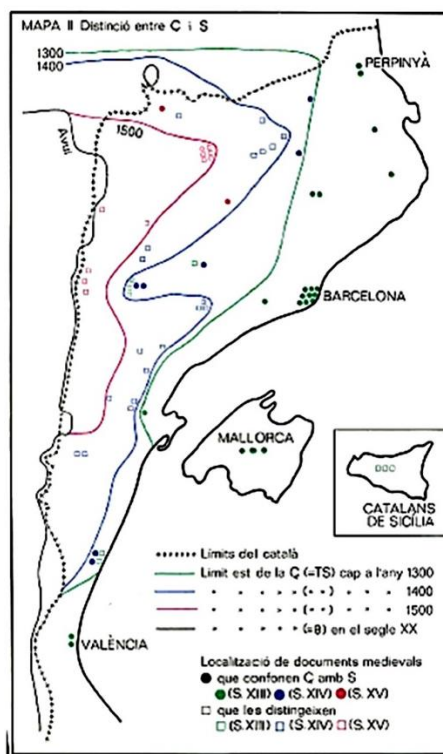


Figura 15. Mapa de distinción entre ç y s (Corominas 1976)

En la documentación del siglo XIV de la Franja de Aragón comprobamos que se mantiene con firmeza la distinción entre ç (*c+e, i*) / *ss* (*s-*, *cons.+s*, *-s*), porque apenas hay ejemplos de confusión, de lo que podemos deducir que en esa época todavía se diferenciaba entre [f̄s] y [s], tal y como lo planteó Coromines en su estudio; estaríamos ante un comportamiento semejante al que él mismo constata en documentación de Morella y Castellón de la Plana (véase Coromines 1976: 22-25). Sería interesante, pues, analizar textos del siglo XV de la Franja para corroborar si en esa centuria se manifiesta ya la confusión de ambos sonidos, tal y como se ha puesto de manifiesto en manuscritos de Castellón de la Plana y de Morella, para observar si, por el contrario, todavía hay distinción en los procedentes de la Ribagorza y del Bajo Aragón, lo cual nos llevaría a poder demostrar que en esas áreas se preservó la articulación de [f̄s], evolucionando posteriormente a [θ] como en aragonés y castellano.

En relación con este aspecto fonético-fonológico se encuentra otro de carácter morfológico: la desinencia verbal de 5.^a persona, que en catalán es actualmente *-u* [-w] (*canteu*) y en aragonés *-z* /-θ/ (*cantaz*). A partir del latín *-TIS*, se observa que el catalán antiguo mantuvo en la escritura durante un tiempo la solución *-ts*, hasta que en el siglo XV comienza a imponerse progresivamente el alomorfo actual, aunque es muy probable que la evolución a [-w] se cumpliera en la lengua popular durante el primer cuarto del siglo XIII (Pérez-Saldanya 1998: 122). En los documentos de la Franja advertimos el mismo conservadurismo que acabamos de mencionar en el catalán general, pero vemos una diferencia interesante entre los pergaminos del siglo XIV de la Ribagorza y los del Matarraña: en los primeros, solo hallamos un caso de vocalización (*aureu*), puesto que lo general es hallar la desinencia *-ts* (*aurets*, *aviats*, *ayats*, *façats*, *fòstets*, *poguéssets*, *poseyscats*, *pregats*, *puxats*, *siats*, *tingats*, *volrets*), en la mayor parte de las ocasiones escrita con *-ç* o *-çs* (*ayaç*, *éaç*, *faç*, *façaç*, *feç*, *forçaç*, *fóseç*, *pageç*, *poreç*, *posaç*,

pregaç, prestaç, puscuaç, puxaç, rogaç, siaç, soç, tengaç, volreç) o *-çs* (*aguésseçs, aureçs, avíaçs, ayaçs, elegiscaçs, enpeyoraçs, espleyteçs ~ spleyteçs, façs, façaçs ~ fàyaçs, fareçs, fósseçs, poguésseçs, possoyreçs, puxaçs ~ puyxayçs, rebaçs, retingaçs, siaçs, soçs, tengaçs ~ tingaçs, tindreçs, volíaçs*), lo cual reafirma, además, la idea de que la grafía ç tenía el valor fonético de [ʧ]. En cambio, en los segundos, aunque también se percibe ese conservadurismo gráfico de *-ts*, lo cierto es que ya comienzan a abundar los casos con *-u* (*aveu, entregeu, espleyteu, leveu, meneu, pageu, penyoreu, tingau, venau, siau, façau, ajau, puxau, vulau, poseyscau, requireu, trobareu, voldreu, volíeu*). Si enlazamos este comportamiento de los documentos con la realidad lingüística actual de la Ribagorza y el Matarraña, veremos que allí donde los documentos reflejan una persistencia del alomorfo *-ts*, hoy encontramos la solución aragonesa *-z /-θ/*, y que allí donde se atestigua un número nutrido de casos con *-u*, hoy hallamos la misma solución catalana. Es evidente, pues, que en los textos ribagorzanos se pone de manifiesto una tendencia a conservar el grupo *-ts* por más tiempo, independientemente de lo que ocurre en otras áreas del catalán y tal vez por influjo del aragonés, lo cual permite explicar el resultado actual de la desinencia, hecho que, por otra parte, indudablemente debe relacionarse con el comportamiento de [ʧ] en la Ribagorza. En este sentido, sería interesante examinar la documentación del siglo XV de esta misma zona, con el fin de verificar si la tendencia conservadora, tal y como la hemos expresado, se mantuvo también durante esa centuria, lo cual sería un testimonio determinante para justificar el alomorfo verbal *-z /-θ/*. Por lo que respecta a la confusión de las consonantes antiguas *yl / ll*, retomamos el estudio citado de Coromines (1976: 17-19), en el que se nos afirma que el sonido [ʎ], cuando proviene de *-LL-* o de una *L-* latina (o prerromana o germánica) en posición inicial, se representa en bastantes textos medievales simplemente con *ll* (incluso, a menudo, sencilla).

Pero, cuando viene de los grupos de yod *-LY-*, *-C'L-*, *-G'L-*, en esos mismos textos el sonido está representado por *yl* (y variantes). Al parecer, en su origen estos dígrafos tuvieron una articulación distinta, que todavía durante los siglos XII y XIII abarcaba la mayor parte del Principado de Cataluña; pero en las centurias siguientes la isoglosa fue retrocediendo hacia el este, hasta limitarse en la actualidad a un área que cubre parte de las provincias de Barcelona y Gerona, donde todavía se mantiene la antigua diferenciación entre [ʎ] y [j] (*aquella* < ECCU(M)ILLA / *aguia* < ACUC(U)LA (véase Figura 16).

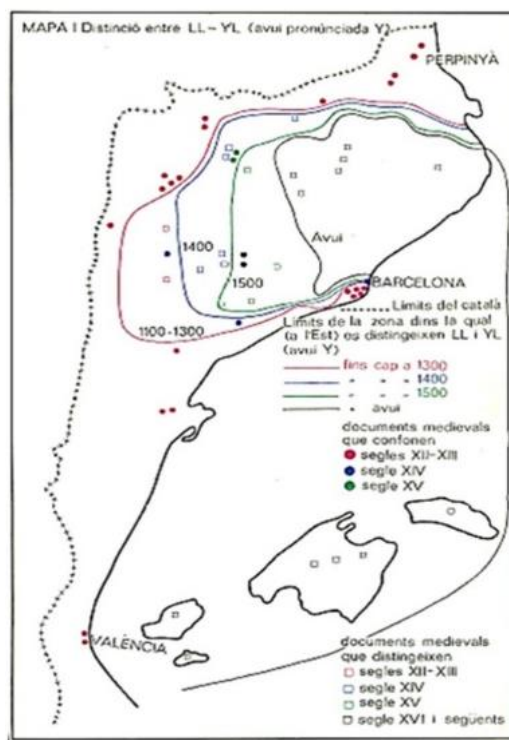


Figura. 16. Mapa de distinción entre -l- y -ll- (Corominas 1976)

En esta cuestión, los documentos de la Franja del siglo XIV también resultan interesantes, puesto que demuestran unas tendencias completamente opuestas de unas zonas a otras. En los documentos ribagorzanos, en el 80 % de los casos, se atestigua la distinción gráfica tal y como la plantea Coromines en su estudio, de manera que aparecen *davall*, *collida* frente a *muyller*, *payller*, *consseyll*, *oveyllas*, *huyll*, *abeyllas*; es decir, no hay confusión gráfica. En cambio, en los textos del Matarraña y el Bajo Cinca se reduce la distinción al 15 % y 5 % de los casos, respectivamente, por lo que se aprecia un predominio del dígrafo *ll* y, en consecuencia, confusión gráfica, como se refleja en las voces *muller*, *palliça*, *perills*, *consell*, confusión que también se observa a la inversa en *davayll*, *eyll*. No obstante, siguiendo las reflexiones de Coromines (1976: 18-19), aunque apenas haya confusión gráfica, en la Ribagorza no podemos pensar en un mantenimiento de la antigua distinción fónica del catalán, porque a lo largo del límite de esta lengua con el aragonés pronto se difuminaron ambos matices confluyendo en uno solo, el de la consonante palatal lateral, del mismo modo que sucedió en aragonés. En consecuencia, lo que testimoniamos en los documentos ribagorzanos es, de nuevo, su conservadurismo, aunque en este caso sea únicamente en el nivel grafemático.

3.2.2. Morfología y sintaxis

Uno de los objetivos que nos marcamos para el estudio de esta documentación era intentar comprender de qué manera se producían los cambios morfológicos y sintácticos a través de las coordenadas diacrónica y diatópica, para intentar confirmar si realmente en aquella sincronía había constituyentes marcados diatópicamente y si se establecían diferencias entre el área septentrional y la meridional de la Franja de Aragón como ocurre en la actualidad.

Desde una perspectiva morfológica, hemos podido constatar ciertas soluciones dialectales comunes entre los textos de la Ribagorza y el Matarraña. Un ejemplo de ello sería el uso del demostrativo *aquestos*, que únicamente se mantiene en en la localidad de Nonaspe (Zaragoza) en convivencia con *estos* (*ALDC VIII*, mapa 1908), y que surge como solución para facilitar la articulación de *aquestes* (DECat III, 718a): *et de aquestos solvem en deffenem a vós e a la universitat de Fraga per tots temps* (Roda, s. XV). En documentación más antigua, se atestigua especialmente en textos occidentales y sólo a partir del siglo XV, al igual que en nuestros manuscritos (CICA; véase también Giralt 2012: 65).

Hemos comprobado, asimismo, que hay un uso sistemático del infijo *-esc-* en los verbos incoativos de la tercera conjugación (más numerosos en los protocolos), general en la lengua hasta finales del siglo XIV: *poseesquen* (Matarraña), *tribuescaç* (Roda), *deffenesch*, *ellesch*, *stituesch* (Roda), *stabllesch*, *constituesch*, *substituesch*, *defenesch*, *elegesch*, *cedesch*, *aconseguesque*, *partesquen*, *poseesque*, *regeesque*, *serveesque* (Matarraña).

Por otra parte, se registran soluciones diferenciadoras entre unos documentos y otros, que responden también a esa realidad viva del catalán del norte oscense y el sur turolense. Un ejemplo es el pronombre átono *lis*, forma de tercera persona del plural con función de CI que se ha constituido analógicamente sobre el singular *li*. Esta forma se anota en los protocolos de Roda: *als quals lis avia signat dia ha dir-lis en la sua qüestió dels linàs; ni lis gosa deffensar los bous*. Ciertamente es que no se atestigua en los pergaminos ribagorzanos del siglo XIV y esporádicamente en los protocolos del siglo XV; pero, aunque prevalezca *los*, su presencia delata que en aquella época debía ser ya de uso corriente en la lengua oral ribagorzana (su empleo en minutas del siglo XVI de La Litera; Giralt 2012: 91). En cambio, en los textos del Matarraña no aparece, del mismo modo que en el catalán actual de esa comarca.

En los documentos de Roda es casi sistemática la terminación *-as* para las formas de femenino plural (*demanar ab altrás cartas et ab altrás obligacions; a las cosas damont ditas present fuy; clams o quereyllas exposar; puxan metre totas aquellas clàusulas*), mientras que en el Matarraña se utiliza *-es* (*en preu de les dites casses; memòria de les dites coses; enteniment de bones persones*). Este comportamiento vocálico también se observa en determinadas desinencias verbales, de manera que tenemos *-a* y *-an* en la Ribagorza (*affermava, sia, era, devia, àgan, pagarian, aurian, afrontan*), pero *-e* y *-en* en el Matarraña (*affermave, sie, ere, devie, agen, pagarien, aurién, afronten*). Si analizamos la realidad dialectal actual, veremos que, en efecto, una buena parte de las hablas catalanas de la Ribagorza (especialmente aquellas en las que se refleja con mayor intensidad la transición lingüística hacia el aragonés) mantienen la A final átona latina como /a/ en todos los casos citados (Veny y Massanell 2015: 259), mientras que en el Matarraña, de acuerdo con el catalán común, siempre como /e/. Por tanto, lo que nos ofrecen los documentos de Roda bien puede ser un reflejo de una solución oral ya fijada en aquella época, y no tanto una tendencia a mantener esa vocal por latinismo, explicación que sí se aduce en los pocos casos hallados en la documentación medieval catalana (véase Giralt y Moret 2018: 172).

Hemos podido testimoniar en los manuscritos del Matarraña las desinencias *-o* y *-on* (de tercera y sexta persona, respectivamente) del presente de subjuntivo de los verbos de la primera conjugación, solución que nunca se registra en los textos de Roda (siempre *-e* y

-en): *çeso, levo, s'atorgo, s'entrego, senyalo, sobro, torno, guànyon, púgon, sàlvon, aprofiton, vàlgon, réston, dónon, pàgon, pàguon, tórnon*. Este rasgo es propio del catalán del Matarraña, además de otros puntos del ámbito noroccidental (Mequinenza, Nonaspe, etc. o localidades de Lleida) y del valenciano septentrional; según los estudios realizados, se documenta por primera vez justamente en estos pergaminos del sur de la Franja de Aragón (Giralt y Moret 2018: 202; Giralt 2019: 38).

En los protocolos de Roda se anotan esporádicamente los posesivos *nostro* (*nostro capítol de Roda*), *vostro* y *vostros* (*dret vostro e dels vostros*), frente a los del catalán común *nostre, vostre, vostres*, que son, a su vez, los usuales en todos los documentos que hemos manejado. Son soluciones dialectales en las que no cabe pensar en una restitución vocálica (como en el caso de las formas catalanas con *-e-*), sino en la conservación del vocalismo etimológico latino; estas son las únicas que se conocen hoy en aquellas hablas de La Litera en las que se refleja en mayor grado la transición hacia el aragonés (Giralt 1998: 123).

Asimismo, en los pergaminos y protocolos de Roda encontramos *hé* como variante de la tercera persona del singular del verbo *ser*, aunque solo en estos ejemplos: *congregats en la nostra casa de capítol hon hé* ('es') *costumat en lo dit monestir; qui ara hé* ('es') *ho per temps serà*. Precisamente esta es la forma que hoy se conserva en el catalán de la Ribagorza y La Litera, paralela a la aragonesa *ye* (Haensch 1960: 119; Giralt 1998: 223), que es la general en un documento de Roda en el que, a diferencia de los documentos anteriores, se trata de un a forma plenamente aragonesa y viva en la actualidad en esta lengua (Moret y Tomás 2014).

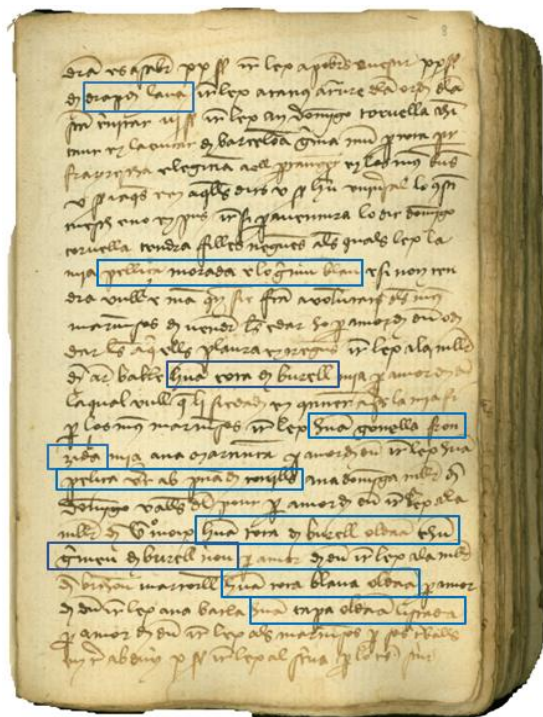
Por último, hay que comentar en los textos de Roda el uso de la preposición *en* 'con' con el valor de la catalana *amb* denota contigüidad o proximidad: *afronta en vinya de Bernat de la Guardiola et en vinya de Arnau de la Guardiola et en vinya de Domingo de Beranuy et en vinya de Domingo de las Poadoras; afronta d'una part en vinya de Ramon Palyarès e de segona part en vinya de Domingo de n'Alegre, terça part en tera de Mateu de Coma; afronta en torrent*. Solamente se ha localizado en el contexto que se observa en los ejemplos, porque en todos los demás casos se utiliza la forma antigua *ab*, la única recopilada en los documentos del Matarraña. En la actualidad, sin embargo, es un uso generalizado en los subdialectos *ribagorçà, tortosí* y *valencià* (Veny y Massanell 2015: 284).

3.2.3. Léxico

Más evidente es aún el interés que este tipo de documentación puede aportar a los estudios que analizan el léxico. Pensar que el lenguaje notarial se restringe a una exclusividad técnica es desconocer qué esconden estos materiales ya que, dada su naturaleza, y como ya hemos ido insistiendo, constituyen una fuente inagotable a la hora de configurar la terminología de una lengua.

La riqueza léxica que albergan las relaciones de bienes que un testador deja a sus herederos, por ejemplo, permiten al investigador arrojar más datos sobre vocablos que no han sido incluidos en los diccionarios más generales o en los grandes bancos de datos que disponemos en la actualidad; o en el caso de que sí lo hayan hecho, se puede matizar su significado, delimitar o adelantar su datación, o determinar su uso compartido entre el aragonés y el catalán. Trabajar sobre terminología, o sobre un

campo semántico concreto, nos posibilita adentrarnos un poco más en el vocabulario usado en la vida cotidiana de una época determinada, como es el caso del ejemplo que presentamos a continuación, un testamento del que pueden extraerse abundantes voces sobre la indumentaria de la época (Giralt 2019):



[...] és a saber XX sous. Ítem lex a pobres a vestir XX sous de **drap de lana**. Ítem lex a catius a renre de la orde de la Senta Trinitat, III sous. Ítem lex a-n Domingo Toruella, abitant en la çitutat de Barcelona, girmà meu, per tota part frayrescha e legítima a ell pertanyent en los meus béns V sous jaqueses, e en aquells dits V sous hereu universal lo constituesch e no en pus. Ítem si per aventura lo dit Domingo Toruella tendrà filles negunes, a les quals lex la mia **pellixa morada** e lo **grimeu blau**, e si no-n tendrà, vull e man que-n sie feta a voluntats del meus marmesós de vendre-les e dar-ho per amor de déu o de dar-les a qui ells plaurà entregues. Ítem lex a la muller de n'Arnau Batle huna **cota de burell** mia per amor Déu, la qual vull que li sie dada en continent après la mia fi per los meus marmesós. Ítem lex huna **gonella fronzida** mia a na Martineta per amor de Déu. Ítem lex huna **pelixa vert ab pena de conills** a na Dominga, muller d'en Domingo Vallès del Pont, per amor de Déu. Ítem lex a la muller d'en Guiamó Moix una **cota de burell oldana** e un **grimeu de burell nou** per amor de Déu. Ítem lex a la muller d'en Bertholomeu Martorell huna **cota blava oldana** per amor de Déu. Ítem lex a na Batla huna **capa oldana listada** per amor de Déu. Ítem lex als marmesós per sos treballs entre abduy X sous. Ítem lex al scrivà per lo testament scriure [...]

Figura 17. Folio 8r del protocolo de Domingo Thomàs del año 1398, conservado en el Archivo Municipal de Fuentespalda, en el que se puede leer el testamento de doña Clara, casera de don Arnau Ripoll, presbítero, habitante de Fuentespalda

En el texto presentado (véase Figura 17), aparecen la *cota*, un vestido que iba desde el cuello hasta aproximadamente los pies, seguramente ceñido y ajustado al cuerpo, usado tanto por hombres como por mujeres; la *gonella*, pieza que llevaban hombres y mujeres, compuesta de cuerpo y falda más o menos larga, generalmente sin mangas, de piel o de tejido más o menos rico; el *grimeu*, especie de túnica larga hasta los pies con aberturas laterales destinadas a los brazos; y la *pellixa/pelixa*, un tipo de abrigo rústico hecho con piel de cabra o de oveja, usado indistintamente por hombres y mujeres. Y de cada una de las prendas de vestir, se dan detalles, como el color (*morada*, *blau/blava*, *vert*), la materia prima (*lana*, *burell* ‘tela tosca de lana e hilo, de color oscuro’, *pena de conills* ‘piel de conejo destinada a forrar una pieza de ropa’), los ornamentos (*listada*, *fronzida*) o la calidad (*nou*, *oldana* ‘vieja’).

En nuestro caso, debemos remarcar que son los protocolos notariales los que nos ofrecen la oportunidad de aproximarnos con más detalle a la realidad material de los pobladores de la Ribagorza, la Litera y el Matarraña (las zonas de donde proceden los documentos notariales que hemos explorado con más detenimiento durante estos años). Estos, por tanto, nos han permitido arrojar luz sobre la adscripción de algunos vocablos a una lengua determinada, a una variedad concreta u observar de qué manera se ha

producido influencia de una lengua sobre otra, en este caso del aragonés sobre el catalán; incluso nos ha permitido documentar determinadas voces que no aparecían en recopilaciones léxicas previas (Giralt y Moret 2019).

Asimismo, estos manuscritos pueden permitirnos situar geográficamente el uso de varias voces que hacen referencia a un mismo concepto, como sería el caso, por ejemplo, de aquellas que aluden al concepto de linde o frontera. Su distribución geográfica durante los siglos XIV-XV es variada, teniendo en cuenta los testimonios estudiados. Con el significado de ‘mojón’, localizamos *molló* (*mullyó*, *muylló*, *mullons*, *muylons*, *muyllons*) y *fito* en documentación de Mequinenza (véase Figura 18), pero *boga* en la documentación procedente de Roda de Isábena (véase Figura 19). Incluso hemos podido registrar por primera vez el sustantivo *boguer* ‘persona que vigila la boga’ (véase Figura 19) o el verbo *abogar*, con el sentido de ‘poner mojones, hitos’ (véase Figura 21): «de aquells secans e regadius e lochs coneguts e per conèxer, abogats e per abogar» (Roda, 1434). En el caso de *abogar*, cabe señalar que se trata de una variante del catalán noroccidental *bogar* ‘afrontar’ (DECat II, 33a), con un significado diferente que coincide con el de *boguear*, solución usada en La Litera (Giralt 2012) y todavía viva en La Codoñera (Teruel) (ALDC IV, mapa 748). Como sinónimo de *abogar* atestiguamos también el verbo (*a*)*mollonar* (*muyllonat*, *amuyllonada* o *amuyllonam*) en pergaminos de Mequinenza (véase Figura 20).

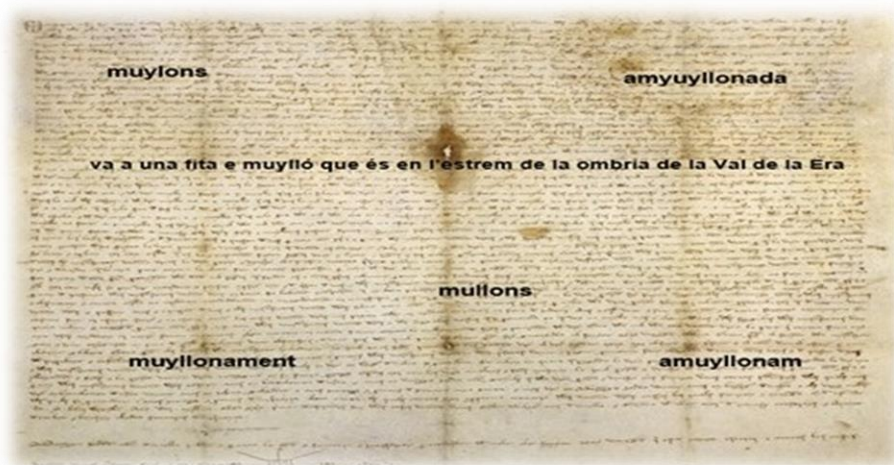


Figura 18. 1331, 23 de noviembre. Mequinenza. AMM. Perg. 50. Not. Domingo Jordà

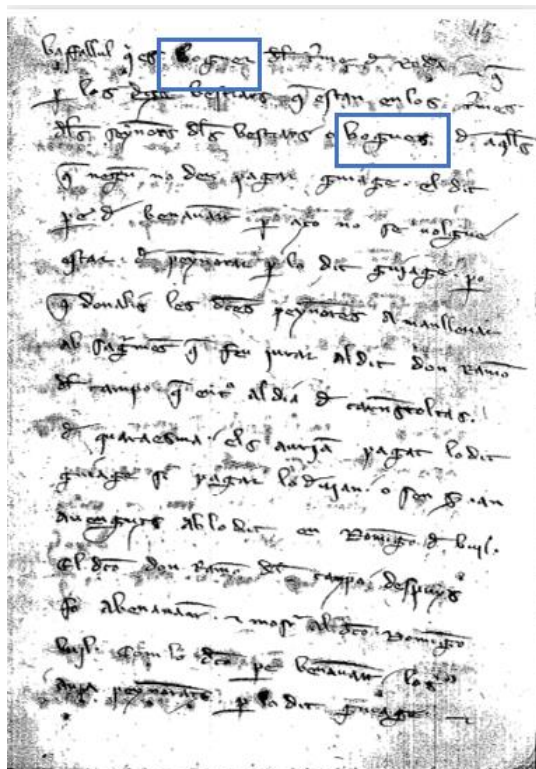


Figura 19. 1316-1319. Roda. ACL. Fondo de Roda. Not. García Sánchez Puértolas: *boguer* /*bogues*

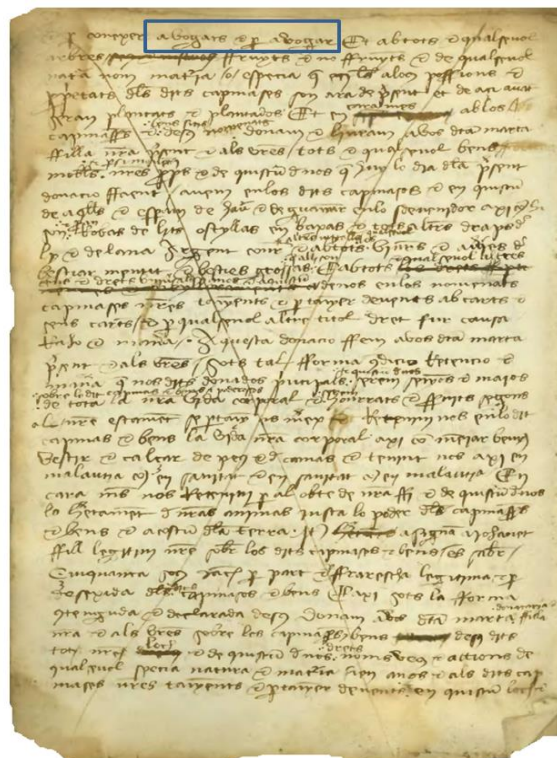


Figura 20. 1434. Roda. AMR. APRoda. Not. Ramon de Santa Creu: *abogats* o *per abogara*

A lo largo de estas páginas hemos pretendido dar cuenta de cómo es posible llevar a cabo un estudio lingüístico basado en documentación antigua de archivo. Tras la localización, edición y análisis de un corpus variado de textos, podemos afirmar que la documentación medieval y moderna conservada hasta el momento y redactada en catalán en poblaciones aragonesas, constituye un material de extraordinario interés para el estudio histórico en general de esta lengua, además de permitirnos testimoniar y estudiar cómo era el catalán escrito en los territorios que en la actualidad han venido a denominarse Franja de Aragón.

Los manuscritos medievales de la zona catalanohablante de Aragón tienen un valor considerable para la filología catalana por diversas razones: (i) porque se trata de textos redactados en el área periférica occidental del ámbito catalanohablante; (ii) porque fueron escritos en un momento en que la lengua catalana tenía ya una fuerte presencia en la administración y en la literatura desarrollada en la Corona de Aragón; (iii) porque, a pesar de haberse concebido en tierras administrativamente aragonesas, nos demuestra que entonces se hablaba catalán en estos territorios; y (iv) porque permiten reafirmar que en esta zona el catalán ha sido la lengua propia hasta la actualidad.

El estudio de los manuscritos reunidos hasta el momento nos ha permitido corroborar que existe una fijación narrativa sometida a unas reglas estrictas y unos hábitos adquiridos, con un formulismo al cual se ha de ajustar cada texto según su contenido; sin embargo, también hemos constatado que, a pesar de presentar una temática poco variada y una ausencia de diferentes niveles del lenguaje, su análisis permite localizar

una serie de elementos que nos acercan a la realidad dialectal de esta zona de Aragón y sus alrededores en aquella época. Este hecho se hace más evidente cuando comprobamos que las coincidencias son bastante numerosas con otros documentos en lengua catalana procedentes de zonas limítrofes, especialmente valencianas o tortosinas. Aunque sea con cierta precaución, podemos poner de manifiesto que los documentos de archivo presentan a menudo una convivencia entre escritura y oralidad, de manera que la primera se manifiesta especialmente a través de los usos arcaicos y de los cultismos que aparecen en cualquiera de los niveles lingüísticos analizados, mientras que la segunda aparece en algunos hechos más concretos, como son la eliminación de grafías no articuladas, la aparición de ciertas alteraciones y evoluciones fonéticas, el uso de determinados rasgos verbales occidentales o noroccidentales y, sobre todo, la utilización de algunos tipos léxicos o variantes léxicas adscritas dialectalmente al área occidental del catalán; sin olvidar, además, los binomios sinonímicos, usados con frecuencia con la intención de aclarar voces cultas, aunque en la mayoría de las ocasiones se trata de un recurso que permite dar énfasis a aquello que se está redactando.

Hasta la actualidad, hemos logrado localizar, editar y estudiar documentación en formato pergamino correspondiente a todo el territorio catalanófono de Aragón durante el período comprendido entre 1304 y 1400; asimismo, existen ya resultados relativos a la documentación en papel, en concreto sobre protocolos notariales de los siglos XIV, XV y XVI procedentes de las áreas de la Ribagorza (Huesca), La Litera (Huesca) y el Matarraña (Teruel); y, finalmente, se ha iniciado el estudio de la documentación del Bajo Cinca (en la zona central de la Franja) redactada durante la segunda mitad del siglo XV y los primeros años del siglo XVI, a partir de la cual podremos ofrecer nuevos datos sobre la realidad de la lengua catalana en la documentación administrativa aragonesa, prestando especial atención a los reflejos diatópicos.

4. Referencias

- ALDC = Veny, Joan / Pons i Griera, Lúdia 2001-2018: *Atles lingüístic del domini català*, Barcelona: Institut d'Estudis Catalans, <<https://aldc.espais.iec.cat/>> [27.02.2020].
- Castillo, Mónica; Díez del Corral, Elena. 2019. *Reescribiendo la historia de la lengua española a partir de la edición de documentos*. Bern-Berlin-Bruxelles-New York-Oxford: Peter Lang.
- CHARTA= *Corpus Hispánico y Americano en la Red: Textos Antiguos*, [http:// www.corpuscharta.es](http://www.corpuscharta.es) [04.05.2021].
- CICA = Torruella, Joan / Pérez Saldanya, Manuel / Martines, Josep (dirs.). *Corpus Informatitzat del Català Antic*, <http://www.cica.cat/index.php> [04.05.2021].
- Coromines, Joan. 1976. *Entre dos llengüatges*. Barcelona: Curial.
- DARA = Gobierno de Aragón (ed.). *Documentos y Archivos de Aragón*. <<http://dara.aragon.es/opac/app/simple>> [04.05.2021].
- DCVB = Alcover, Antoni M.; Moll, Francesc de Borja. *Diccionari català-valencià-balear* <http://dcvb.iecat.net/default.asp/> [03.05.2021].
- DECat = Coromines, Joan 1980-2001. *Diccionari etimològic i complementari de la llengua catalana*, 10 vols. Barcelona: Curial Edicions catalanes.
- García-Miguel, José M. 2022. Lingüística de corpus: de los datos textuales a la teoría lingüística. *Estudios de Lingüística del Español* 45: 11-42.

- García-Valle, Adela. 2004. Las fórmulas jurídicas medievales: un acercamiento preliminar desde la documentación notarial de Navarra. *Anuario de historia del derecho español* 74: 613-640.
- Giralt, Javier. 1998. *Aspectos gramaticales de las hablas de La Litera (Huesca)*. Zaragoza: Institución «Fernando el Católico».
- Giralt, Javier 2010. «Llits de roba» del siglo XVI en Albelda (Huesca). En Rosa M.^a Castañer y Vicente Lagüéns, coords. *De moneda nunca usada: estudios filológicos dedicados a José M^a Enguita Utrilla*. Zaragoza: Institución «Fernando el Católico», pp. 295-310.
- Giralt, Javier. 2012. *La llengua catalana en documentació notarial del segle XVI d'Albelda (Osca)*. Ayuntamiento de Albelda.
- Giralt, Javier. 2017. La Franja de Aragón: un vínculo lingüístico indefectible con Cataluña. En A. Sabio Alcutén, coord. *Tejidos de vecindad. Los vínculos históricos entre Aragón y Cataluña, siglos XVIII-XX*. Zaragoza: Diputación de Zaragoza/Gobierno de Aragón, pp. 509-527.
- Giralt, Javier. 2019. *Llibre de testaments (1398-1429) d'en Pere Oriola, notari de Fondespata. Edició i estudi lingüístic*. Teruel: Instituto de Estudios Turolenses.
- Giralt, Javier; Moret, María Teresa. 2018. «*Sie manifesta cosa a tots hòmens*». *El català del segle XIV en textos notariais del Matarranya (Terol)*. Zaragoza: Servicio de Publicaciones. Universidad de Zaragoza.
- Giralt, Javier; Moret, María Teresa. 2019. Manuscrits medievals del Matarranya (Terol) i llurs aportacions al lèxic català. *Caplletra* 67 (Tardor): 39-64.
- Giralt, Javier; Moret, María Teresa. 2021. Manuscrits de la Ribagorça vs. Manuscrits del Matarranya. *Zeitschrift für Katalanistik* 34: 227-247.
- Goebel, Hans. 1975. Qu'est-ce que la scriptologie?. *Medioevo Romano* II: 3-43.
- Haensch, Günther. 1960. *Las hablas de la Alta Ribagorza*. Zaragoza: Institución «Fernando el Católico».
- Ibarretxe-Antuñano, Iraide; Hijazo-Gascón, Alberto; Moret-Oliver, María Teresa. 2017. The importance of minority languages in motion event typology: The case of Aragonese and Catalan. En I. Ibarretxe-Antuñano, ed. *Motion and Space across languages and applications*. Amsternarm/Philadelphia: John Benjamins Publishing Company, pp. 123-149.
- Monjo, Marta. 2004. *Sarrains sota el domini feudal. La Baronia d'Aitona al segle XV*. Lleida: Servei de publicacions de la UdL.
- Moret, Hèctor; Sorolla, Hugo. 2018. La llengua catalana en la documentació administrativa de Fraga a la baixa edat mitjana i moderna: de la pervivència a la substitució. En M. Pérez Saldanya y R. Roca Ricart, eds. *Del manuscrit a la paraula digital / From Manuscript to Digital Word: Estudis de llengua i literatura catalanes / Studies of Catalan language and literature*. Amsterdam / Philadelphia: John Benjamins, pp. 50-66.
- Moret, María Teresa. 2010. *Documentació notarial aragonesa del segle XIV escrita en català. Edició i estudi grafemàtic*. Zaragoza: Tesis doctoral inédita.
- Moret, María Teresa. 2020. El catalán en textos medievales aragoneses no literarios. En J. Giralt y F. Nagore, eds. *Aragonés y catalán en la historia lingüística de Aragón*. Zaragoza: Prensas de la Universidad de Zaragoza, pp. 124-149.
- Moret, María Teresa; Tomás, Guillermo. 2014. *El pleito del guaje ganadero de Ribagorza (1316-1319). Edición y estudio lingüístico*. Zaragoza: Fuentes

- Históricas Aragonesas 70. Institución «Fernando el Católico» (C.S.I.C.). Excma. Diputación de Zaragoza.
- Moret, María Teresa; Giralt, Javier. 2021. Léxico de la indumentaria en testamentos de los siglos XIV y XV de Fuentespalda (Teruel). *Archivo de Filología Aragonesa* 75: pp. 41-69.
- Moret, María Teresa; Ibarretxe-Antuñano, Iraide; Hijazo-Gascón, Alberto. 2017. Tipología semántica i llengües minoritàries. Els esdeveniments de moviment en les llengües d'Aragó. En J. Giralt y M. T. Moret, eds. *El repte d'investigar sobre la Franja d'Aragó. Jornada de l'Associació Internacional de Llengua i Literatura Catalanes a Saragossa (28 d'octubre de 2016)*. Zaragoza: Servicio de Publicaciones de la Universidad de Zaragoza, pp. 197-207.
- Moret, María Teresa; Ibarretxe-Antuñano, Iraide. 2019. Una aproximación desde la tipología a la deixis de (ir y venir) en aragonés i catalán medievales. En J. M^a Enguita; M^a A. Martín-Zorraquino; V. Lagüéns y M^a L. Arnal, eds. *Actas del X Congreso Internacional de Historia de la Lengua Española (Zaragoza, 7-11 de septiembre de 2015)*. Zaragoza: Institución «Fernando el Católico», pp. 2145-2158.
- Pagarolas, Lluís. 1999. Els arxius històrics. *Bibliodoc. Anuari de Biblioteconomia, Documentació i Informació*: 155-166.
- Rasico, Philip. 2006. *El català antic*. Girona: Universitat de Girona - CCG Edicions.
- Riesco, Ángel. 2003. *Vocabulario Científico-Técnico de Paleografía, Diplomática y Ciencias afines*. Madrid: Barrero-Azedo Ediciones.
- Sánchez-Prieto, Pedro. 2012. Desarrollo y explotación del corpus de documentos españoles anteriores a 1700 (CODEA). *Scriptum Digital* 1: 5-35.
- Santanach, Joan. 2012. Les edicions d'Els Nostres Clàssics i els criteris d'edició. Aportació a la historia de la col·lecció. *Anuari Verdaguer* 20: 261-292.
- Tomás, Guillermo. 2020. *El aragonés medieval. Lengua y Estado en el reino de Aragón*. Zaragoza: Prensas de la Universidad de Zaragoza.
- Veny, Joan; Massanell, Mar. 2015. *Dialectologia catalana. Aproximació pràctica als parlars catalans*. Barcelona: Publicacions de la Universitat de Barcelona – Publicacions de la Universitat de València – Publicacions de la Universitat d'Alacant.

¹ Una publicación imprescindible que abraza el panorama actual de las ediciones filológicas y el aprovechamiento de los textos para los estudios lingüísticos es Castillo y Díez del Corral 2019.

² Algunos autores han planteado la posibilidad de que este sonido hubiese existido también en el catalán preliterario general (Rasico 2006: 180-181); sin embargo, Rasico (2006: 196-197) concluye que, a partir de las grafías, difícilmente se puede saber si realmente fue así, y que un cambio de /ts/ a /θ/ y posteriormente a /s/ no solo hubiese sido muy excepcional, sino poco o nada natural desde un punto de vista fonológico. Estamos completamente de acuerdo con este autor y consideramos que es más sensato pensar que, tanto en las hablas ribagorzanas de transición como en las bajoaragonesas, el grupo /ts/ siguió una evolución paralela a la del aragonés, incluso tal vez por influjo de esta lengua. No hay que olvidar, en este sentido, que en el catalán de transición ribagorzano se conserva /-θ/ como alomorfo de la desinencia verbal de la 2.^a persona del plural, como en aragonés, frente al catalán general /-u/, hecho morfológico del que nos ocupamos a continuación.

Trabajo de campo *

Iraide Ibarretxe-Antuñano
Universidad de Zaragoza – IPH
iraide@unizar.es

Resumen

El trabajo de campo consiste en la recogida sistemática de información sobre un determinado fenómeno *in situ*, en su entorno natural. En este artículo, tras una breve revisión sobre el lenguaje y la diversidad lingüística, se exponen las bases que sustentan a esta metodología lingüística. Este artículo se divide en dos partes fundamentales. En primer lugar, se explican las diferentes concepciones que abarca esta metodología en el área de la lingüística y los requisitos necesarios para llevar a cabo la interacción entre el investigador, la comunidad lingüística y la investigación. En segundo lugar, se enumeran los pasos necesarios para realizar un trabajo de campo y se profundiza en algunos aspectos prácticos y tecnológicos. Finalmente, a modo de conclusión, se reflexiona sobre la importancia y la necesidad de llevar a cabo trabajo de campo en las ciencias del lenguaje.

Palabras clave: trabajo de campo en lingüística, datos empíricos, comunidad lingüística, diversidad lingüística, metodología lingüística.

Abstract

Fieldwork can be described as the systematic data gathering about a specific phenomenon that takes place on its natural environment. After a short introduction about languages and linguistic diversity, this paper introduces the basics of fieldwork methodologies. It starts with an overview about the concept of fieldwork in Linguistics, followed by a description of the main scientific and ethical requirements needed to conduct fieldwork around language. Then, it proposes a ‘must-to-do’ list of basic steps to bear in mind in fieldwork, with a special focus on practical and technological issues. The final part discusses the importance, the scope, and the need to incorporate fieldwork in the language sciences.

Keywords: Linguistic fieldwork, empirical data, linguistic community, linguistic diversity, linguistic methodology.

1. Introducción: el ser humano, el lenguaje y las lenguas

El lenguaje es una de las características más especiales que tiene el ser humano y que, al menos, de momento, nos diferencia de otras especies animales. A través del lenguaje, los seres humanos nos comunicamos con otras personas y, por supuesto, con nosotros mismos. Gracias al lenguaje, las personas somos capaces de pensar y crear pensamientos. A fin de cuentas, el lenguaje es eso, comunicación y cognición. A los seres humanos se nos supone que nacemos preparados, o al menos predispuestos, físicamente, para producir lenguaje. Todos tenemos el mismo cuerpo, el mismo cerebelo y, en teoría, estando sanos, no hay nada que nos impida a ninguno de nosotros desarrollar el lenguaje. Primero,

gestualmente, y poco a poco, en diferentes fases, gestual y oralmente; porque, no hay que olvidarse de que el lenguaje es multimodal, es decir, no solamente se procesa de forma vocal-auditiva –como a veces parece que nos hacen creer– sino también de forma visuo-gestual (para una revisión, véase Ibarretxe-Antuñano y Valenzuela Manzanares (2021: cap. 5)). Ahora bien, el lenguaje, por mucho sustrato biológico que se le presuponga, no estará nunca completo, ni se desarrollará de forma adecuada, sin la parte socio-cultural. Los seres humanos son animales sociales, crecen y se desarrollan de forma grupal, no aislada, y el lenguaje, como no podía ser de otra manera, también lo es. Descritas ya algunas de las características básicas del lenguaje humano según las investigaciones más recientes, es decir, que es multimodal y biocultural, es momento para centrarse en cómo se manifiesta el lenguaje humano. En otras palabras, es momento de describir las lenguas y la diversidad lingüística.

Actualmente, existen más de 7000 lenguas, tanto orales como signadas, distribuidas por todo el mundo. A la existencia de lenguas diferentes y su distribución espacial por el mundo es lo que se conoce con el nombre de diversidad lingüística. Aunque parezca sencillo a simple vista, acotar la diversidad lingüística no es tarea baladí, debido a la conjunción de diversas razones de distinta naturaleza. Para empezar, el número de lenguas exacto no está claro, puesto que habría que empezar por definir y, sobre todo, acotar dónde termina una variedad y empieza una lengua; una tarea nada fácil porque su interpretación depende, entre otros factores, que muchas veces no son lingüísticos ni científicos, del contexto lingüístico en el que se hablen esas lenguas o variedades.

A este factor, se le une, además, el hecho de que las lenguas o las variedades no conocen de límites jurídico-administrativos, sino que suelen expandirse, más bien, a través de diferentes extensiones geográficas, independientes de las fronteras políticas que, como se sabe, suelen ir cambiando de lugar a lo largo de la historia (véase, por ejemplo, el caso ilustrativo del euskara en Aurrekoetxea 2022). Estos factores pueden explicar, al menos parcialmente¹, por qué cuanto mayor sea la diversidad lingüística, más tendencia hay de agrupar variedades bajo una misma lengua general (p. ej., hablar de “chino” (como lengua general) y de “dialectos chinos” a pesar de ser estos oralmente ininteligibles entre sí (véase Ramsey 1987)), y viceversa, es decir, cuanto menor es la diversidad lingüística, más se tiende a buscar “lenguas” y ver las variedades como más dispares (p. ej., referirse al “sueco” o al “danés” como lenguas distintas a pesar de ser mutuamente inteligibles).

A esta situación, compleja ya de por sí desde el punto de vista estrictamente filogenético y tipológico, se le unen además toda una serie de situaciones sociolingüísticas que suelen tenerse en consideración en las caracterizaciones generales de estas lenguas. Estos factores sociolingüísticos serían, entre otros, la existencia de un sistema de escritura (y, por ende, de documentación escrita), la creación de una variedad estandarizada –que no hay que olvidar que siempre es artificial (se construye intencionalmente) y prescriptiva (uso obligado de determinados rasgos lingüísticos) (Ibarretxe-Antuñano 2021a: 58)– y el estatus de oficialidad de estas lenguas. Estos factores, que suelen tenerse en cuenta para estudiar, por ejemplo, la vitalidad de las lenguas (Unesco 2003), suelen reflejar muy poco la verdadera realidad de la mayoría de las lenguas del mundo –la mayoría ni están escritas, ni estandarizadas, ni son oficiales– sino más bien, suelen ser el espejo de la situación sociolingüística de las lenguas que se consideran mayoritarias (principalmente, desde el prisma occidental, es decir, lenguas como el inglés, el francés o el español) (véanse Moreno Cabrera 1997; Amorós Negre 2014; Ibarretxe-Antuñano 2021a, b).

Sea como fuere, lo que está claro de este tipo de inventarios de lenguas es que, como se aprecia en la Figura 1: (i) hay una gran diversidad de lenguas “variadas”, visualmente representadas en el mapa a través de diversos colores y formas, y (ii) están distribuidas de forma irregular por el mundo.

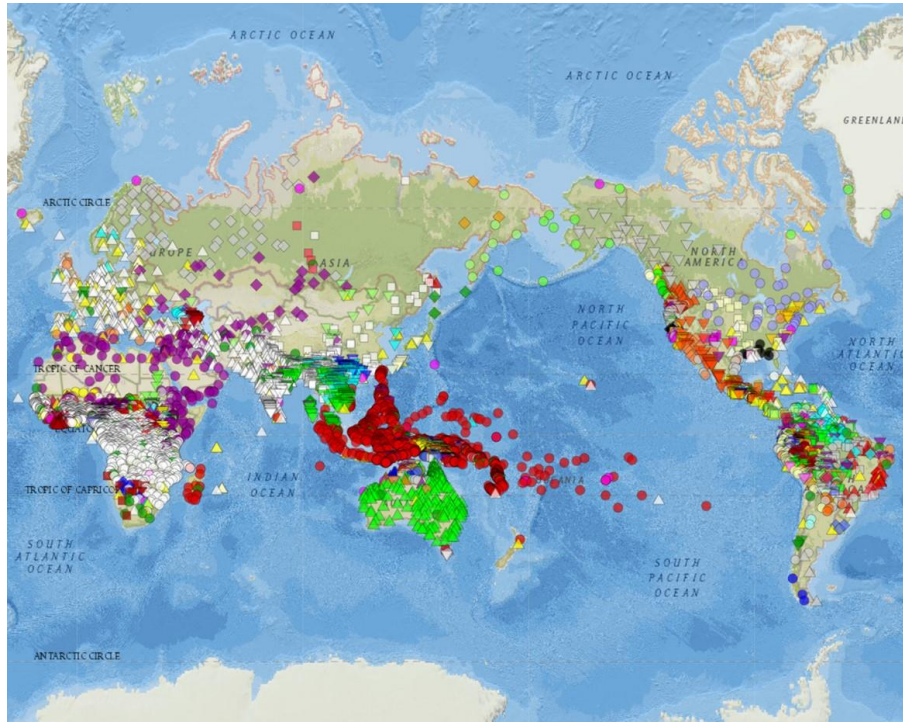


Figura 1. Diversidad lingüística. Mapa creado con Glottogob 4.5.
[<https://glottolog.org/glottolog/language>]

Las zonas a ambos lados del Ecuador son las que, en líneas generales, concentran el mayor número de lenguas. En términos de países y continentes, Europa destaca por la poca diversidad lingüística, en comparación con otros continentes como Oceanía, África y las Américas. Y, los países europeos, con mayor diversidad lingüística de la que a veces nos han hecho creer (un país → una lengua), siguen estando claramente en desventaja con respecto a países como, por ejemplo, Papúa Nueva Guinea (que ostenta el récord de mayor diversidad lingüística con 864 lenguas (18 extintas)) o Nigeria (con 533, con 23 extintas) o México (298, en 63 nacionales) (para mayor información, se puede consultar el Ethnologue o la plataforma de libre acceso Glottolog 4.5; véase sección 6).

En las imágenes que aparecen en la Figura 2, se aprecia claramente cómo, en términos de diversidad lingüística, el tamaño no importa. En esta imagen adaptada de Harrison (2007), se aprecia cómo lugares con poca extensión geográfica terrestre, como Oceanía, se vuelven extensos en términos de diversidad lingüística. En los mapas de la Figura 2, está marcado el archipiélago de Vanuatu como un ejemplo ilustrativo. Apenas se aprecia en el mapa de la Figura 2.a., pero destaca en el de la Figura 2.b.



Figura 2.a. Extensión geográfica

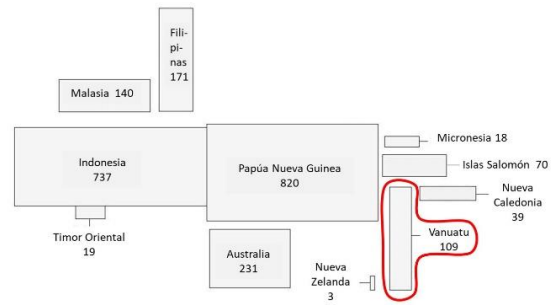


Figura 2.b. Diversidad lingüística

Figura 2. Oceanía: tan pequeño y tan grande. Mapas adaptados de NatGeo Map Maker y Harrison (2007: 13)

Estas diferencias en cuanto a la distribución de la diversidad lingüística nos dan aún más pistas sobre el ser humano y su relación con el lenguaje y las lenguas. El estado natural del ser humano no es ser monolingüe, sino plurilingüe (más del 60 % de la población mundial lo es desde que nacen, aunque, de todas esas lenguas, el 96 % de la población habla solo un 4 % de ellas). Y, las lenguas, como manifestaciones concretas del lenguaje, son el mayor recurso que tiene el ser humano para cristalizar y transmitir la cosmovisión de sus hablantes ya que, como se menciona en el informe de la Unesco (2003: 2), recopilan “saberes únicos, culturales, históricos y ecológicos”. Por eso, el hecho de que una lengua deje de tener hablantes, al igual que cuando se extingue una especie animal o vegetal, no es solamente una terrible pérdida para el patrimonio de la humanidad. Además del desarraigo que pueda causar en los descendientes de esos últimos hablantes, cada lengua es crucial tanto para entender cómo funciona el lenguaje humano como para entender “la prehistoria humana y el mantenimiento de los diversos ecosistemas del mundo” (Unesco 2003: 3).

Cada lengua guarda la sabiduría de cada pueblo y si se pierde, no solo se pierden “palabras” sino todo el saber que estas encierran puesto que han sido creadas para transmitir un saber específico que solo puede interpretarse desde un entorno específico. En un reciente artículo sobre la relación entre medicina, plantas y lenguaje, Cámara-Leret y Bascompte (2021) mostraron que el conocimiento sobre los beneficios de ciertas plantas medicinales se encuentra íntimamente relacionado con su expresión lingüística en una lengua determinada. Es decir, la existencia de las plantas medicinales puede salir fuera de esa comunidad lingüística, pero el conocimiento sobre aspectos concretos de su uso y su beneficio están circunscritos a la transmisión de los mismos en esas lenguas en concreto. Estos autores, a través de estudio de las plantas medicinales en tres zonas conocidas por su diversidad ecológica y lingüística, América del Norte, noroeste de la Amazonia y Papúa Nueva Guinea, concluyen que el conocimiento sobre las especies botánicas está en peligro pero que su protección no reside exclusivamente en el amparo de estos especímenes. El primer paso para cuidar este conocimiento medicinal, de hecho, está en la propia documentación y protección de las lenguas en las que se expresa.

Este ejemplo de íntima simbiosis entre el lenguaje y la botánica ilustra muy bien la importancia del trabajo de campo y de cómo llevarlo a cabo. Cuando se hace trabajo de campo en lingüística, no solo se está investigando la morfosintaxis o la fonética de una lengua, sino también su significado y su uso por parte de sus hablantes; hablantes, por otra parte, que no están desconectados del lugar, la sociedad y la cultura en la que desarrollan sus vidas diarias. Como se verá a continuación, el trabajo de campo lleva

consigo esa “inmersión” *in situ* en el entorno de la lengua que se está investigando, es parte del ADN de este tipo de metodología lingüística; el tipo, el tiempo y la forma de “inmersión” se ajustarán a las necesidades de la investigación que se ha de llevar a cabo, pero, siempre ha de estar presente. En la siguiente sección se hace una revisión por el propio concepto de trabajo de campo, las diferentes perspectivas que hay sobre qué puede ser considerado el trabajo de campo y los elementos básicos de esta metodología.

2. ¿Qué es el trabajo de campo?

Uno de los pilares de cualquier estudio científico son los datos, es decir, los materiales sobre los cuales se construye y se sustenta una teoría sobre un determinado fenómeno. Da igual el área de conocimiento en la que se trabaje, lingüístico como en este artículo, o el objetivo final de la investigación, la documentación de una lengua desconocida o la descripción del funcionamiento de un rasgo sintáctico; hoy en día, cualquier método científico incluye una parte empírica, de manejo de datos, y una parte teórica, de manejo de hipótesis y explicaciones. Al fin y al cabo, son las dos caras de una misma moneda: la investigación científica.

Dentro de las posibles formas de las que disponemos para desarrollar una investigación se encuentra lo que se conoce como el **trabajo de campo**, es decir, la recogida sistemática de información sobre un determinado fenómeno *in situ* y de primera mano. Esta primera definición está descrita, a propósito, de manera muy general, ya que se puede aplicar a cualquier campo científico, ya sea la lingüística, la arqueología, la química o la biología, pero recoge dos aspectos fundamentales del trabajo de campo que se irán perfilando de forma más específica en las siguientes páginas. Por un lado, la investigación *in situ*. Es decir, el investigador o el equipo de investigadores es el que se desplaza de su lugar habitual de trabajo (su laboratorio, su oficina...) al sitio donde, de forma natural, se encuentra el fenómeno que se quiere estudiar. Este desplazamiento implica no solamente un desembolso económico importante (p. ej., viaje, estancia, materiales de investigación), sino también un esfuerzo personal ingente (p. ej., tiempo, adaptación). Por otro lado, la recogida sistemática de información sobre un fenómeno, es decir, el objetivo es recabar información de primera mano a través de diferentes técnicas previamente diseñadas y reflexionadas sobre un aspecto concreto. En otras palabras, antes de realizar una campaña, es necesario saber y tener previsto de antemano qué se va a investigar, por qué se va a investigar y cómo se va a llevar a cabo la investigación. Cualquier metodología científica lleva en su preparación estas tres preguntas, pero, en el caso del trabajo de campo, se vuelven aún más importantes porque, generalmente, las campañas se llevan a cabo en momentos puntuales y por un tiempo determinado, por lo que hay que estar preparado con antelación para cualquier incidencia. En otras palabras, el trabajo de campo no es coger una maleta e ir a la aventura a ver qué se encuentra, sino que requiere una preparación que a veces excede con creces las tareas propias de un investigador.

Como ya se ha mencionado, esta definición de trabajo de campo puede servir para cualquier área de conocimiento, pero en este artículo se va a centrar en el trabajo de campo relacionado con el lenguaje. Así que se van a perfilar en más detalle algunos de los aspectos que se han introducido en esta definición general del trabajo de campo: el sitio de investigación, el investigador y lo investigado.

2.1. El sitio de investigación

Se acaba de mencionar que uno de los requisitos del trabajo de campo es el desplazamiento por parte del investigador al lugar donde se produce de forma natural el fenómeno que se está investigando. En el caso del trabajo de campo en lingüística esto implica desplazarse al sitio donde se encuentran hablantes de una determinada variedad lingüística. Aunque parezca sencillo definir el trabajo de campo, existen diferentes opiniones entre los especialistas sobre lo que se puede considerar trabajo de campo debido a la localización de esas comunidades de hablantes. Posiblemente, una de las ideas románticas que se puede tener sobre el trabajo de campo es precisamente el del lingüista que viaja a lugares “exóticos” y remotos –léase todo aquello ajeno a lo occidental– a trabajar en una comunidad cuya variedad lingüística está aún sin describir científicamente. Por eso, si se hiciera un pequeño test sobre las fotos que se incluyen en la Figura 3 y se preguntara cuál de ellas corresponde al trabajo de campo, posiblemente se escogerían las de la Figura 3.a. en lugar de las de la Figura 3.b.



Figura 3.a. Isla Kaile’una
Islas Trobiand, Papúa Nueva Guinea

Figura 3.b. Ansó
Pirineo occidental, Aragón, España

Figura 3. Trabajo de campo diverso

Desde el punto de vista occidental, las imágenes de la Figura 3.a. pueden representar lo que se tiene en el imaginario colectivo con respecto al trabajo de campo: un lugar lejano donde la comunidad en la que se trabaja, sus costumbres, su comida y su ecosistema difieren radicalmente del mundo occidental y cuya lengua, normalmente en peligro de extinción, o al menos, amenazada, no ha sido descrita suficientemente o tiene alguna característica que la hace especial. Las fotos de la Figura 3.a., salvando los estereotipos que se acaban de describir, encajan con este tipo de contexto. Estas instantáneas son del lingüista Günter Senft, uno de los mayores especialistas en la lengua Kilivila. En las fotos se ve su lugar de trabajo de campo, la aldea Tauwema en la Isla Kaile’una (Islas Trobiand, Papúa Nueva Guinea), y una sesión con uno de sus colaboradores, Mokeilobu. Las fotos de la Figura 3.b. corresponden al trabajo de campo de mi grupo de investigación en la

villa de Ansó, en el Pirineo occidental aragonés, donde llevábamos a cabo una sesión sobre la expresión del movimiento en aragonés ansotano con la Sra. Pilar Mendiara. Ambos lugares de trabajo cumplen con la definición del trabajo de campo que se ha descrito más arriba y que suele asociarse a este tipo de metodología –o incluso, a esta rama de la lingüística– y que, en parte, tiene como objetivo principal la documentación y la descripción de una lengua de estas características. Sin embargo, hoy en día también se considera trabajo de campo el realizado en los entornos representados en la Figura 4.



Figura 4.a. Casa



Figura 4.b. Escuela



Figura 4.c. Museo

Figura 4. Trabajo de campo “urbano”

Estas imágenes no están tomadas ni en lugares remotos, ni sus investigadores han pasado una larga temporada buceando en una cultura diferente, ni tenían como objetivo documentar una lengua menos explorada. La Figura 4.a. corresponde a una sesión de elicitación de datos sobre gestualidad con dos hablantes de español en su casa en Zaragoza (Aragón). La Figura 4.b. muestra una instantánea de otra recogida de datos con niños bilingües euskara-español en la Ikastola Astileku (Portugalete, Bizkaia). Finalmente, la Figura 4.c. está tomada de un festival de divulgación científica en un museo de Utrecht (Países Bajos) donde los investigadores recogían datos basados en un juego de comunicación artificial (Dingemanse et al. 2014). Todas las imágenes de la Figura 4 representan trabajo de campo que, a veces, ha recibido, con mayor o menor aprecio, la etiqueta de “urbano” por realizarse en ciudades o en entornos industrializados. Para autores como Crowley (2007) o Hyman (2001), este tipo de investigación no sería precisamente trabajo de campo por no cumplir las premisas arriba mencionadas. Sin embargo, en estos casos, lo que es cierto también es que es el investigador, y no el informante, el que se desplaza al entorno natural donde se encuentran los informantes y el que se adapta a esa comunidad lingüística para cumplir con las tareas que necesitan ser llevadas a cabo. Al igual que en las otras situaciones de la Figura 3, en este trabajo de campo urbano y “cercano” –denominado así por ser el entorno que más se puede asemejar al del investigador o el que no requiere un desplazamiento a otro continente o región– también surge una colaboración entre el investigador y la comunidad lingüística y es, a raíz de esa colaboración y en ese contexto, donde se recoge y se analizan los datos lingüísticos. Meakins et al. (2018) mencionan que este tipo de trabajo de campo urbano suelen llevarlo a cabo investigadores que no tienen los medios para desplazarse o que quieren investigar lenguas de herencia; sin embargo, esto no es así necesariamente; todo depende del objetivo del estudio y del tipo de informantes (lengua, variedad, edad, contexto...) que se necesite. Por ello, la investigación que se realiza en una comunidad lingüística cercana y urbana también se ha de considerar trabajo de campo. Chelliah y de Reuse (2011), proponen, de hecho, la distinción entre los términos de “lingüista en inmersión” y “lingüista de entrevista” para distinguir, pero, al mismo tiempo, incluir todo tipo de trabajo de campo. El primer término describiría al investigador de campo prototípico, el que pasa largas temporadas en convivencia con la comunidad lingüística,

generalmente, no urbana y remota, y se integra en el día a día de su comunidad. El segundo sería el investigador que mantiene una relación más superficial y corta, que se circunscribe a las interacciones para las sesiones de trabajo con los integrantes de la comunidad. Como bien señalan Chelliah y de Reuse (2011: 7-9), cada tipo tiene su propia caracterización, sus pros y contras y, en muchos casos, la calidad de trabajo no depende del tipo de trabajo de campo en sí mismo, sino la efectividad con la que se diseña y se lleva a cabo (véase sección 3). A fin de cuentas, para realizar cualquier tipo de trabajo de campo, sea urbano, de inmersión, remoto o de entrevista, es indispensable cumplir con los mismos requisitos y los mismos pasos que se describen en las siguientes secciones. Eso sí, habrá que adaptarlos a las características y a las necesidades específicas de cada investigación. En este artículo se va a adoptar una visión comprensiva del trabajo de campo, de ahí que se incluyan ejemplos procedentes de diversas tipologías, siempre y cuando, los datos recopilados no procedan de la introspección (de la intuición del propio lingüista como hablante nativo) e incluyan la interacción con los miembros de la comunidad en sus entornos naturales.

2.2. El investigador y “lo investigado”

El **investigador** o lingüista, puesto que en este artículo nos estamos refiriendo a la labor de este tipo de científico, es aquella “persona” que lleva a cabo el trabajo de campo; es decir, que se encarga de preparar la campaña (véase sección 3.1.), de llevar a cabo la recogida y el análisis de los datos y de estar en colaboración con la comunidad lingüística. La palabra “persona” está entrecomillada a propósito, puesto que, en la mayoría de las situaciones, el trabajo de campo es un trabajo en equipo. No solo porque organizar una campaña de trabajo conlleva la movilización de diferentes agentes, lo que requiere colaborar con otros colegas de profesión para preparar materiales, conocer información sobre la lengua, etc., sino también porque, en muchas ocasiones, para poder interpretar bien los datos, se necesita la colaboración de otros profesionales (antropólogos, sociólogos, historiadores, filólogos...) así como la mediación de especialistas dentro de la propia comunidad. En este sentido, se suelen distinguir tres tipos de “lingüista”, lo que puede ser útil para darse cuenta de que el trabajo de campo es accesible a cualquier investigador, pero que conlleva la existencia de ciertas ventajas y desventajas, así como diferentes grados de adaptación y de riesgo de incurrir en lo que Labov (1972) denominaba la “paradoja del observador”, es decir, el efecto o la influencia que puede tener la presencia del investigador en el comportamiento y la producción lingüística de sus informantes (véanse Dwyer 2006; Meakins et al. 2018).

El primer tipo es el lingüista “interno” (*insider*). Este es el profesional que pertenece a la comunidad lingüística en la que se realiza la investigación; y, por lo tanto, no solamente ya conoce la situación y el contexto lingüístico, sino que además tiene la ventaja de hablar la lengua y tener acceso a la comunidad. Normalmente, suele estar más implicado en el proceso ya que es su propia comunidad la que se está estudiando, aunque, al formar parte de la comunidad, el investigador puede llegar a alienarse del resto, esto es, sus propios paisanos pueden considerarle externo a su propio grupo. El segundo tipo es el lingüista “externo” (*outsider*); el profesional que viene de un contexto lingüístico diferente y que puede abarcar desde casos en los que ha de aprender la lengua (p. ej., cuando un investigador europeo, hablante de alemán, aprende e investiga kilivila, una lengua austronésica de Papúa Nueva Guinea), hasta casos en los que, aun conociéndola, procede de un entorno diferente (ciudad-campo). Este tipo de lingüista tiene la desventaja obvia

de tener que hacerse camino en una comunidad que no es la suya, en la que siempre puede ser el “outsider”. Sin embargo, una vez superado el periodo de adaptación, puede tener la ventaja de estar en una posición idónea para ser más receptivo a las singularidades de la lengua en la que trabaja, para poder influir positivamente en el empoderamiento de la comunidad lingüística (“alguien de fuera se interesa por nosotros”) y para poder atraer el interés de fuera de la comunidad sobre ella misma (divulgación, subvenciones, etc.). El tercer tipo, el lingüista “interno-externo” (*insider-outsider*), hace referencia al profesional que, sin ser de la misma comunidad lingüística, procede de un contexto similar (social, cultural, lingüístico, etc.), por lo que puede ser más sensible a ciertos aspectos relacionados con la lengua y la comunidad lingüística (p. ej., un hablante de lengua minoritaria, como el euskara, aprende e investiga aragonés, otra lengua minoritaria).

A pesar de estas diferencias, cualquier investigador de campo que vaya a estudiar una comunidad lingüística, la suya, una ajena o una similar, ha de cumplir, o al menos ha de tener presente, que el trabajo de campo es siempre “**en colaboración**” con la comunidad lingüística, por lo que debe tener claro **qué va a dar, qué va a recibir y cómo lo va a usar** (véase Austin y Sallabank 2011). Generalmente el lingüista de campo suele ser consciente de que siempre, en mayor o menor medida, ha de encontrar formas para apoyar a esa comunidad lingüística a través de su investigación y su presencia (Evans 2008). Estas formas son múltiples: desde incluir una nota de agradecimiento a su comunidad en las publicaciones científicas o enviarles copias de estos trabajos científicos, hasta divulgar sobre sus resultados de investigación en cualquier foro no profesional y transferir sus conocimientos en acciones concretas dentro de la propia comunidad.

Además, quizás de manera aún más crucial, el lingüista ha de mostrar **respeto** por la comunidad en la que convive, por sus deseos y por sus costumbres. Ha de ser consciente, por ejemplo, de que quizás no estén de acuerdo con ser fotografiados, con responder a ciertas preguntas o con llevar a cabo ciertas tareas. Ha de comprometerse con los acuerdos de uso y disfrute de los datos a los que ha llegado con la comunidad (véase sección 3.1.1.2.1.). Ha de integrarse en la comunidad, no solo porque esta le está ofreciendo uno de sus mayores tesoros, su lengua; sino porque para entenderla va a necesitar hacerse con toda la cosmovisión en la que se incardina esa lengua y sus hablantes. Esa “integración”, como se ilustra en la Figura 5, puede pasar por tener que tirarse por el suelo, jugar con muñecos, pintarse la cara, participar en las actividades programadas en un colegio, hasta por establecer una tradición cada vez que se está en el campo como, por ejemplo, organizar e invitar a una comida a todos los miembros de la comunidad con los que se ha convivido o también aprender a preparar la comida típica de dicha comunidad e iniciarse en sus rituales propios.

En la Figura 5, se ilustran estos tres tipos de “integraciones” de distinto cariz. La Figura 5.a. muestra una fiesta infantil (celebración de una final del equipo de fútbol local) en la Ikastola Astileku (Portugalete, Bizkaia) donde la investigadora pasó de recoger datos con niños bilingües euskara-castellano, a hacer las veces de “fotógrafa-reportera” de una fiesta escolar. La Figura 5.b. recoge la comida en un restaurante de Felipe Carrillo Puerto (Yucatán, México) con la que el lingüista Juergen Bohnemeyer, especialista en maya yucateco, agasaja y despide a sus colaboradores al final de su temporada de trabajo de campo en Yaxley. La Figura 5.c. muestra a la lingüista Vera da Silva-Sinha, especialista en lenguas tupi-guaraní (hatxa kuñ, awetý, kamaiurá), con la pintura corporal ceremonial y preparando la comida tradicional *beiju* de estas comunidades de la reserva Xingú en Brasil.



Figura 5.a. Celebrando fiestas (arriba izquierda) Figura 5.c. Pinturas y comidas
Figura 5.b. Festines de despedida (abajo izquierda) y festines tradicionales (centro, derecha)
Figura 5. Las “otras tareas” del investigador en el trabajo de campo

Otro ingrediente fundamental de cualquier lingüista de campo es la **empatía** lingüística, es decir, “la capacidad de identificarse con alguien y compartir sus sentimientos y comprobar que lo raro, lo malo y lo gracioso en las lenguas es una cuestión de perspectiva” (Ibarretxe-Antuñano 2020: 72). En otras palabras, un lingüista (en realidad, cualquier persona) jamás debería pensar que la lengua o la variedad de otra persona está mal, debería de buscar por qué se usa y a qué fenómeno lingüístico corresponde. Un lingüista nunca debería atosigar ni forzar la respuesta que necesita para corroborar su hipótesis, ni tampoco debería proyectar su perspectiva o su intuición lingüística para corregir la interpretación del dato lingüístico que su informante le está ofreciendo. A veces, aunque se esté trabajando con lenguas/culturas parecidas, no se debe dar por hecho que sus estructuras y sus interpretaciones han de ser también parecidas. Por mencionar una idea análoga que puede ilustrar este error, es pensar en los “falsos amigos” que estudiamos cuando aprendemos una segunda lengua: superficialmente iguales, pero profundamente diferentes.

Ahora bien, junto a esta empatía, es necesario mantener un equilibrio justo entre el interés científico genuino por una lengua diferente, con características “sorprendentes”, y el peligro de que la “sorpresa gozosa” se convierta en una actitud paternalista por la cual se pase a tratar, aunque sea de forma inconsciente, a la comunidad lingüística y a su lengua como piezas de museo, en vez de como lenguas vivas, modernas y adaptadas para la comunicación de cualquier aspecto del mundo.

Por último, aunque solo por el orden de discusión de este artículo, es necesario comentar más en detalle el concepto de **comunidad lingüística**. En este trabajo, este término ha aparecido ya en numerosas ocasiones. Se ha utilizado para hacer referencia de forma global y comprensiva al grupo de individuos que integran una “unidad” que comparte una misma lengua o variedad y un mismo entorno sociocultural. De nuevo, esta definición es amplia y genérica a propósito; ha de cubrir todas las comunidades lingüísticas en las que se puede llevar a cabo una campaña de recogida de datos. Es decir, desde una

comunidad como los trobriandeses o los awetý, hasta un grupo de hablantes seleccionados por una franja de edad, un nivel de formación, una variedad lingüística, entre otros.

Además de considerar la comunidad lingüística como un todo en el que el investigador ha de integrarse, las relaciones que se establecen con los miembros de estas comunidades pueden ser de diferentes tipos; lo que también hace que existan, en la bibliografía especializada, diferentes formas de referirse a los integrantes de estas comunidades como, por ejemplo, como “colaboradores”, “informantes”, “participantes”, “hablantes”, entre otras. Cada uno de estos términos lleva consigo una carga de significado adicional: “colaborador” incide en que el trabajo de campo es tanto del investigador como de la comunidad lingüística que participa y comparte su saber; “informante” puede resultar más neutra, pero posiciona a la comunidad como un ente pasivo que se limita a transmitir información; “participante” es aún más aséptica y hace referencia a que dicha persona toma parte en una tarea, por lo que esta denominación incluye también a los que participan en tareas de laboratorios, es decir, fuera de esos entornos naturales. Finalmente, “hablante” se suele usar cuando se hace referencia al hecho de que esta persona tiene competencias lingüísticas en la lengua de estudio, aunque, como se describe a continuación, no hay que olvidar que se suelen distinguir diferentes tipos de hablantes y que formar parte del trabajo de campo de un investigador implica posiblemente mucho más que el ser capaz de comunicarse en una lengua. En definitiva, todos estos términos se usan, a veces, de forma indistinta, pero hay que conocer las diferencias que se pueden hacer en relación a los individuos que integran una comunidad lingüística y su papel en la investigación.

Así, para empezar, se puede hacer una distinción práctica –y gradual– entre el tipo de involucramiento que puede tener alguno de los miembros de la comunidad lingüística con el investigador. Como se explicará en detalle en la sección 3.3.1., muchas veces el acceso para llevar a cabo el trabajo de campo puede mediar a través de un miembro de la comunidad. En algunas ocasiones, esta persona puede ser ya un lingüista interno, es decir, alguien de la propia comunidad que, además de hablante, haya realizado investigaciones al respecto. Por ejemplo, en la Figura 8.e., se muestra una sesión de trabajo en la que participa la investigadora, Vera da Silva-Sinha (mujer a la izquierda), y el lingüista interno, Wary Kamaiurá Sabino (hombre en el centro como camiseta azul). En otras ocasiones, esta persona puede no ser especialista, pero ser capaz de: (i) comunicarse en diferentes lenguas, por ejemplo, la lengua de estudio y otra *lingua franca* de comunicación (no olvidemos que en el trabajo de campo en muchas ocasiones el investigador puede tener unas competencias lingüísticas restringidas en la lengua que estudia), (ii) tener y/o desarrollar “aptitudes” para el trabajo lingüístico, desde llevar a cabo aspectos prácticos, como transcribir los datos, hasta conceptuales, como explicar estructuras y significados, que son cruciales para un desarrollo óptimo de la investigación, y (iii) estar dispuesto a colaborar estrechamente (y pacientemente) con el investigador. Por ejemplo, la Figura 3.a. muestra al lingüista Günter Senft y su colaborador Mokeilobu transcribiendo datos en kilivila. A este tipo de colaborador se le suelen otorgar diferentes nombres según los especialistas: “profesor de lengua”, “asistente de campo” o “ayudante lingüístico” (Crowley 2007: 86). Finalmente, se podría incluir como otro tipo de involucramiento el del grupo de hablantes que toman parte en las tareas propias de elicitación del investigador, pero que no toman parte ni en la mediación con la comunidad ni en las tareas de transcripción o análisis lingüísticos (véase Figura 4).

Otro aspecto que también hay que tener en cuenta con respecto a estos colaboradores es que, según el objeto de estudio y, en muchas ocasiones, según su disponibilidad y el grado de vitalidad de la lengua, el informante puede tener diferentes tipos de contacto o de relación con su lengua. Meakins et al. (2018: 11) distinguen entre tres tipos: “hablante nativo”, el que desde el nacimiento está expuesto y utiliza la lengua en todos los contextos de uso de forma habitual; (ii) “hablante semi-nativo”, el que estuvo expuesto a la lengua desde el nacimiento, pero, por diversos motivos, no la utiliza o no es competente en todos los ámbitos de uso y (iii) “hablante *rememberer*”, el que tiene un conocimiento pasivo de la lengua, pero no la habla. Por razones obvias, a no ser que se haga un estudio específico (p. ej., segundas lenguas, lenguas de herencia, etc.), el candidato óptimo es el primero, aunque no siempre se tiene la suerte de contar con informantes de este tipo. En cualquier caso, como se explica en detalle en la sección 3.1., en todo trabajo de campo se recoge y especifica la información psico-socio-lingüística completa de cada uno de los participantes en las tareas de elicitación (véase cuestionario biográfico de la Figura 7).

3. ¿En qué consiste el trabajo de campo?

Como se ha introducido en la sección anterior, en este artículo, la visión del trabajo de campo es amplia, ya que abarca cualquier tipo de situación en la que se colabore con hablantes de la comunidad lingüística en su entorno natural, al margen del tipo de lengua/variedad o del tipo de campaña que se vaya a realizar (véase Chelliah y de Reuse (2011: 10-24) para una caracterización crítica e informada de diferentes aproximaciones al trabajo de campo). Por eso, independientemente de las particularidades de cada investigación, existen unas bases y fases comunes a cualquier tipo de trabajo de campo que se van a describir a continuación. Dadas las restricciones de este tipo de publicación, solo hay espacio para ofrecer algunas indicaciones a modo de mini-guía para poner en situación al lector interesado en esta metodología; quienes quieran más detalles o información actualizada sobre dichos aspectos pueden consultar obras generales de referencia, así como los recursos que se recopilan en la sección 5.

3.1. Mini-guía básica para el trabajo de campo: el antes, el mientras y el después de la campaña

Como cualquier otra metodología de investigación, el trabajo de campo se puede organizar alrededor de tres fases consecutivas principales que van desde los preparativos iniciales, siguen con la recogida de datos y terminan con la producción derivada de ese análisis y trabajo. Quizás la nota distintiva en el trabajo de campo es que todo se organiza alrededor del momento estrella: la campaña de recogida de datos. En esta sección, se caracterizará cada una de estas fases. Es importante tener en cuenta que cada investigador tiene su propio “manual de campo” y que, dependiendo de las variables que ya se han comentado en las secciones anteriores, algunos de los pasos, requisitos o materiales puede que no se apliquen a todos los tipos de campañas que se pueden realizar. No obstante, a continuación, se describen los detalles esenciales de cada fase basados en la experiencia personal y profesional de la que escribe y que ha sido aprendida a través de colaboraciones con diversos equipos de investigación especializados en este tipo de metodología. Esta perspectiva hace que todos los elementos que se incluyen estén constatados, es decir, se describan desde la experiencia, pero, también, hace que, por ejemplo, a la hora de

ilustrarlos falte (o sobre) algún matiz. Por ello, el lector ha de tener en cuenta que existen otras posibilidades y que lo que se muestra, a continuación, son solo algunos posibles ejemplos. En cualquier caso, los elementos que se describen en esta sección son los primordiales, los que no se pueden olvidar y los que marcan la diferencia entre un simple “tomar a un hablante y pasarle un cuestionario” y un articulado proceso de elicitación en el campo.

3.1.1. Fase 1: la pre-campaña o antes de iniciar la investigación *in situ*

En esta fase se incluyen todos los elementos que han de tenerse en cuenta antes de iniciar el viaje al lugar de la investigación. Posiblemente sea el más costoso en todos los sentidos puesto que se ha de preparar la investigación (qué y cómo se va a investigar), todo el equipo (físico e inmaterial) para poder llevarla a cabo en el campo y la logística de la propia campaña (desplazamiento, estancias, permisos...). El trabajo de campo requiere que el investigador prepare hasta el último detalle de la campaña para que así pueda solucionar cualquier tipo de imprevisto que pueda surgirle mientras lleva a cabo su investigación. Esta preparación no es ajena a otras metodologías de investigación, especialmente a aquellas en las que se depende de la participación y el reclutamiento de personas (por ejemplo, tareas de laboratorio), pero quizás en el trabajo de campo sea aún más crucial puesto que el investigador se desplaza al sitio de investigación, muchas veces alejado del lugar de origen y, además, durante un tiempo significativo y sin los medios necesarios al alcance de la mano como en un laboratorio. Todo esto implica no solamente tiempo y dinero, sino también una gran responsabilidad, ya que el investigador tiene solamente una oportunidad para llevar a cabo esa investigación. Si la pierde, es posible que tenga que esperar a la siguiente campaña o, incluso, que no pueda volver a repetirse. En cualquier caso, si se da, no volverá a hacerlo en las mismas circunstancias. Por ello, esta primera fase es larga y costosa, pero muy necesaria.

3.1.1.1. Decidir la investigación: el qué, el cómo y el porqué de una campaña

El primer paso para preparar una campaña es tener claro qué comunidad lingüística se va a investigar y por qué. Esta decisión se puede tomar por diversos motivos: porque esa variedad no está suficientemente descrita, porque esa lengua tiene un rasgo lingüístico que es crucial para testar alguna hipótesis, etc. Tradicionalmente, se suele distinguir entre dos tipos de perspectivas: la de la **documentación de la lengua**, donde se quiere documentar las prácticas y tradiciones lingüísticas de una comunidad lingüística (Himmelman 1998; Gippert et al. 2006), y la de la **descripción lingüística**, donde se quiere describir las bases de una lengua a partir de textos, o lo que se conoce como la llamada trilogía boasiana: la gramática, el vocabulario y los textos (Epps et al. 2017). En la práctica, ambas perspectivas son complementarias y necesarias; además, en ocasiones el grado de aplicación de una u otra depende de la propia lengua y el objetivo de estudio concreto que se tenga. De cualquier forma, lo que es importante recordar es que, generalmente, cuando se va al campo, se aprovecha la estancia al máximo. Es decir, aunque el objetivo primordial sea la obtención de datos para constatar un fenómeno lingüístico concreto a través de un test específico, sería recomendable ir más allá y recoger información de uso general. Es decir, se puede investigar diferentes aspectos fonético-fonológicos, morfo-sintácticos y semántico-pragmáticos, pero siempre informados desde la recopilación de un “corpus” que, como señalan Meakins et al. (2018: 9), ha de incluir

textos discursivos de diferente tipología (narraciones, elicitaciones, conversaciones, diálogos, anotaciones, definiciones...) así como textos de géneros menos convencionales como los musicales o los artísticos, que pueden mostrar el alcance intertextual y peso cultural de ciertas características lingüísticas de una comunidad (Barwick 2012). De hecho, ese corpus deberá ser multimodal para poder así llevar a cabo un análisis comprensivo de esa lengua o fenómeno lingüístico no solo en su forma oral sino gestual (Seyfeddinipur 2012).

Sea documentación o descripción, lo primero que tiene que hacer el investigador en esta fase previa a la campaña es **recopilar** toda la información posible tanto de la lengua² que va a estudiar (descripción general y explicación de fenómenos lingüísticos específicos) como de la comunidad lingüística en la que va a trabajar. Esta información debe ser extraída no solamente de fuentes bibliográficas, sino también a través del contacto con otros investigadores que hayan trabajado ya en esa comunidad lingüística. A veces puede haber una tendencia (errónea completamente) de que el estudio de una lengua o variedad o de un aspecto lingüístico concreto es “propiedad” de un investigador. Nada más lejos de la realidad. Además de que las lenguas no son de un investigador, el trabajo de campo es necesariamente colaborativo: se necesita trabajar en conjunto con otras personas para la recogida de datos, pero igualmente se necesita contar con el apoyo y el saber de otros investigadores, no solo para los aspectos propiamente lingüísticos, sino también porque pueden ayudar a entender e informar de cómo hay que actuar para integrarse en esa comunidad lingüística. Este aspecto es importante: no se puede ir al campo sin más, aunque sea a unos kilómetros del lugar de trabajo, porque el estudio de una lengua no solamente es la descripción de su estructura lingüística, o el saber, a través de un informante, si se dice (a) o (b) en una variedad; sino que es el estudio de toda una cosmovisión del mundo a través de la experiencia de otras personas. Por este motivo, si se quiere acceder a una comunidad lingüística es necesario conocerla al máximo.

Unida a esta fase de recopilación de información, el siguiente paso es **contactar** con la comunidad lingüística. Esta es posiblemente una de las partes más complicadas para los lingüistas externos puesto que en muchos casos no se tiene acceso directo a la comunidad. Generalmente el contacto se puede establecer de varias maneras: (i) a través de otros investigadores que ya han estado o tienen contacto con esa comunidad o con alguna otra cercana; (ii) a través de centros sociales, asociaciones culturales, órganos administrativos..., es decir, a través de agentes dentro de la propia comunidad que puedan tener interés en su lengua y, al mismo tiempo, tengan acceso a posibles informantes y, finalmente, (iii) a través de lo que aquí se va a llamar “aventura y paciencia”, esto es, conseguir acercarse a la comunidad lingüística, sin previo contacto, e integrarse de forma pausada. Este último método recuerda mucho a algunos de los investigadores pioneros en la exploración de otras lenguas y culturas como, por ejemplo, Franz Boas en Norteamérica, pero también a otros más cercanos que, como Jean-Joseph Saroïhandy, acudían al Pirineo desde otras partes de Europa para estudiar lenguas como el euskara, el aragonés, el occitano o el catalán, o ya en épocas más recientes a Josefa Canellas y su investigación sobre el asturiano. Es una forma complicada y, además, poco recomendada como única forma de acercamiento, pero que inevitablemente va a estar presente en el trabajo de campo, puesto que, en diferentes grados, el investigador de campo ha de estar siempre abierto a nuevas posibilidades (a conocer otros informantes, otras comunidades...) y a tener estoicismo para colaborar con cualquier comunidad lingüística.

3.1.1.2. El equipo necesario para la campaña

Otro paso clave en esta primera fase, además de la preparación de la propia **logística** de la estancia (viaje, alojamiento, documentación (visados, vacunas, etc.)), es la preparación del equipo que se va a llevar al sitio de trabajo. Bajo esta etiqueta genérica de equipo, se va a incluir aquí el **equipo físico**, es decir, equipamiento, materiales de grabación y materiales de elicitación, así como el **equipo inmaterial**, esto es, los permisos y las autorizaciones correspondientes. Aunque el tipo de equipo va a depender de la campaña y ha de adaptarse a las condiciones de sitio en el que se vaya a realizar (temperatura, electricidad, etc.), por regla general, siempre que se pueda, se llevará más material “por si acaso” (p. ej., más tarjetas de memoria, más cargadores, más copias) y “todo listo” (p. ej., fotocopias o materiales impresos). Esto siempre tiene que ser así, aunque pueda darse el caso de tener acceso a estos elementos en el sitio de campaña.

3.1.1.2.1. Equipo inmaterial

No hay que olvidar lo que ya se ha mencionado en estas páginas sobre la improvisación en el trabajo de campo. Es importante tener recursos para poder salir al paso de situaciones sobrevenidas o para complementar lo que ya se tiene organizado (p. ej., buscar nuevos informantes para sustituir otros), pero todo trabajo de campo requiere una preparación minuciosa que a veces lleva meses conseguir. Como, por ejemplo, ocurre cuando el investigador tiene que pedir **autorizaciones** para poder llevar a cabo la investigación. Estas autorizaciones a veces se necesitan para trabajar con una comunidad lingüística (p. ej., en lugares protegidos de comunidades indígenas es necesario pedir permisos a las autoridades competentes), pero también se piden a las propias instituciones de investigación para certificar que la investigación que se llevará a cabo cumple con todas las normas de seguridad y protección. Generalmente, estos centros de investigación tienen su propio “Comité de ética de la investigación” que es quien, previo informe sobre los objetivos y las características de la campaña, aprobará la investigación (y la aplicación y el disfrute de los fondos para llevarla a cabo en la mayoría de los casos).

Además, ya con vistas a la propia colaboración con la comunidad lingüística, el investigador ha de preparar con antelación los documentos de información, autorización y **consentimiento informado** de las sesiones de trabajo para que sean firmadas por los propios individuos de la comunidad que tomen parte en la investigación. La finalidad de estos documentos es múltiple. Por un lado, informan al participante de los objetivos y del tipo de investigación que se va a llevar a cabo, qué se espera de su colaboración y cuáles son las condiciones (p. ej., si recibe una compensación (monetaria o de otra índole) por su participación). Esta explicación ha de redactarse de forma clara y adaptada al informante, es decir, de manera que no se requiera un conocimiento especializado para comprenderla, y generalmente se da a leer o se comunica en voz alta, para que después el informante pase a firmarla. Por otro lado, estos documentos piden explícitamente el consentimiento del informante para que autorice a los investigadores a utilizar estos datos (audio-visuales) para unos fines, en una forma y en unos foros concretos (p. ej., imagen y sonido para investigación y docencia en cursos y congresos). En la Figura 6, se muestran dos ejemplos de autorizaciones.

Departamento de Lingüística General e Hispánica Universidad Zaragoza

MovEs

Departamento de Lingüística General e Hispánica Universidad Zaragoza

MovEs

Portugués, 1 de febrero de 2015

Estimado padre:

Nos gustaría dar las gracias por dejar que vuestra hija participe en la recogida de datos lingüísticos que se va a llevar a cabo en la Basilea Alemana a partir de marzo de 2015 para el subproyecto de investigación "La adquisición temprana de los eventos de movimiento en niños bilingües (español-sukara)". El objetivo principal de este trabajo es la creación de un corpus lingüístico de datos lingüísticos de 15-20 años bilingües en español y sustrato de idiomas comprendidos entre 3,5 y 4,0 hasta la edad de 9,0 años, para su poder investigar su desarrollo lingüístico.

Esta investigación se encuentra dentro del proyecto de investigación "Movimiento y espacio desde la tipología semántica II: Variación y adquisición (MovEaE-VIA)". Este proyecto está financiado por el Ministerio de Economía y Competitividad del Gobierno de España (FF2013-45553-C3-1-P) y dirigido por la Dra. Ibarretxe-Antuñano. Esta investigación se lleva a cabo en el Departamento de Lingüística General e Hispánica de la Universidad de Zaragoza.

Los datos lingüísticos que recogamos en estas sesiones a lo largo de los próximos meses van a ser grabados con cámara de vídeo. La tarea principal que tendrá que realizar vuestra hija es la descripción de algunas imágenes/vídeos ante el investigador tanto en sustrato como en español. Los datos recogidos serán utilizados exclusivamente con fines docentes e investigadores. Los datos recogidos no serán divulgados en foros públicos que queden fuera del ámbito docente e investigador.

Hacéis un cordial saludo,

Iraide Ibarretxe-Antuñano

Dra. IP. Iraide Ibarretxe-Antuñano
Investigadora Principal del Proyecto MovEs
Departamento de Lingüística General e Hispánica
Universidad de Zaragoza

PERMISO DE UTILIZACIÓN DE RESULTADOS DE EXPERIMENTOS LINGÜÍSTICOS PROYECTO MovEs

Fecha: _____

Por la presente, yo, _____, con D.N.I. _____, padre/madre/tutor legal de _____, DOY PERMISO Y AUTORIZO a los investigadores del proyecto "Movimiento y espacio desde la tipología semántica II: Variación y adquisición (MovEaE-VIA)", financiado por el Ministerio de Economía y Competitividad del Gobierno de España (FF2013-45553-C3-1-P) a que dispongan de los datos audiovisuales recogidos en los experimentos lingüísticos en los que mi hija ha participado para fines exclusivamente docentes e investigadores. Los datos recogidos no serán divulgados en foros públicos que queden fuera del ámbito docente e investigador.

Fdo. _____
Padre/madre/tutor legal

Figura 6.a. Proyecto MovEs

e.o.s.s. evos@mpi.nl

Formulario del participante

Esta información se mantiene separada de las respuestas de las tareas. Se usa para confirmar que has entendido las tareas y que el experimentador te recompensará por tu participación. Esta información no se usará para ningún otro propósito.

Si tienes alguna pregunta sobre el experimento sobre cómo se usarán los datos, no dudes en preguntar al experimentador. Puedes, si quieres, pedir el compromiso ético asociado a este Proyecto. También puedes enviarnos un mensaje a evos@mpi.nl o visitar la página web del proyecto <http://www.mpi.nl/evos> para obtener más información.

Gracias por participar.

- Declaro que entiendo las tareas de EoSS tal y como se me han explicado.
- Estoy de acuerdo en llevar a cabo estas tareas.
- Entiendo que se grabará el audio de mis respuestas. Estoy de acuerdo en que estas respuestas sean almacenadas en un repositorio y que mi nombre no será asociado a esas grabaciones de sonido.
- Entiendo que mis datos personales solo se utilizarán de la forma descrita arriba.

NOMBRE/APELLIDOS: _____

FIRMA: _____ FECHA: _____

Por mi participación recibiré la suma de _____

PAGO RECIBIDO (al contado)/ FORMULARIO DE PAGO COMPLETADO (para transferencia bancaria): _____

Figura 6.b. Consorcio EoSS

Figura 6. Autorizaciones y consentimientos de informantes

En la Figura 6.a. se muestra el formulario de consentimiento perteneciente al proyecto MovEs de la Universidad de Zaragoza para una investigación con menores, por lo que se requería la autorización de un tutor. En la Figura 6.b. se muestra otro consentimiento, esta vez de adulto, perteneciente al consorcio de investigación Evolution of Semantic Systems (EoSS) del Max Planck Institute for Psycholinguistics de Países Bajos.

3.1.1.2.2. Equipo físico

Dentro de esta categoría, se va a incluir el equipamiento técnico necesario para poder preparar y documentar las sesiones de trabajo, así como todos los materiales de investigación y elicitación necesarios para llevarlas a cabo.

Con respecto al **equipamiento técnico**, aunque, hoy en día, llevar un dispositivo de grabación digital como, por ejemplo, un móvil, está ya casi al alcance de todos, puede ser más cómodo para la recogida de datos y puede sacarnos del paso en cualquier situación, el trabajo de campo requiere de un equipamiento algo más sofisticado. La razón está clara: el trabajo de campo no se puede tomar solo como un medio para conseguir un beneficio personal, sino para cumplir con unos objetivos de investigación concretos, para los cuales una calidad de grabación específica o el hecho de grabar también la imagen no son irrelevantes. El trabajo de campo es una inversión a largo plazo que puede ser útil para otros investigadores (interesados en otros fenómenos lingüísticos que puedan requerir un buen sonido y una buena imagen) y, desde luego, para la propia comunidad lingüística con la que se está colaborando (creación de materiales didácticos, exposiciones y documentación de la lengua, etc.). A fin de cuentas, el investigador está documentando uno de los patrimonios inmateriales más importantes de la humanidad, sus lenguas y, por lo tanto, es importante que lo haga en las mejores condiciones posibles y con vistas a que sea un legado a posteriori para la humanidad.

Aquí, además de no disponer de espacio para hacer recomendaciones específicas sobre este tipo de equipamiento, entra en juego el hecho de que estos aparatos están en continuo desarrollo. Tampoco se puede obviar que según el sitio de investigación (condiciones atmosféricas, acceso a energía, etc.) habrá que ajustar más o menos el equipo. Aunque las necesidades no son las mismas cuando se trabaja en mitad del Amazonas, de Siberia, o

del Pirineo, como se describe a continuación, los cables, los cargadores, las baterías (incluidas las solares) y las tarjetas de memorias de repuesto nunca sobran ni están de más, aunque haya enchufes, conexión wifi y se tenga un almacenaje en la nube ilimitado. De igual manera, lo que se va a grabar también importa: no es igual una sesión con hablantes de lenguas orales que de lenguas de signos o una en la que las tareas estén dirigidas al estudio de la gestualidad (Seyfeddinipur 2012). En este sentido, aunque las grabaciones suelen estar dirigidas a la obtención de datos, es importante documentar fotográficamente la campaña: la sesión, los informantes (si dan permiso), el sitio de investigación, las costumbres, etc. No hay que olvidar que el trabajo de campo es una metodología colaborativa y que este tipo de documentación de campaña no solo puede ser importante para ilustrar la propia investigación científica (como se ha hecho en este artículo con las imágenes reales que aparecen) sino servir para la propia comunidad lingüística.

Así que, a continuación, más que precisar un tipo de aparato específico, se va a proporcionar una lista de equipamiento que, a ser posible, no debería faltar. Este equipamiento se suele escoger por su robustez, durabilidad (baterías), capacidad de almacenaje y compatibilidad de sistemas. Es también importante no olvidar que todo equipo requiere un desembolso económico importante y que no siempre lo más caro es necesariamente lo mejor; equipos menos sofisticados pueden ser igualmente válidos, solo hay que saber sacarles provecho. En definitiva, el equipamiento sería el siguiente:

- ordenador y complementos (ratón, cables, baterías): además de ser el medio de trabajo del investigador, se suele usar también para el desarrollo de tareas. Es importante tener cargadores de emergencia (especialmente en zonas donde no se dispone de electricidad en el sitio de trabajo).
- equipo de grabación (por orden de preferencia): cámara de video con complementos (formato .avi/.avchd, conectable micro externo y auriculares; trípode, cargador, baterías (2), bolsa/maleta de viaje, cables de conexión); grabadora de audio (mínimo 24bit/96kHz/estéreo/formato .wav; cables, batería); cámara de fotos (cargador, batería, cables de conexión).
- equipo de sonido: micrófono externo, paravientos y auriculares. Aunque los dispositivos de grabación tienen una calidad excelente tanto para conseguir una grabación óptima como para escucharla, es recomendable el uso de estos aparatos.
- equipo de almacenaje: disco duro externo, lápices y tarjetas de memoria (velocidad, compatibilidad y capacidad). Aunque estamos acostumbrados a guardar nuestros documentos en la nube, esta no está disponible ni en todas partes ni en todo momento, por eso es necesario tener este tipo de equipamiento siempre, tanto para guardar las sesiones de trabajo como para guardar los materiales de elicitación, autorización y toda la documentación relativa a la investigación y a la campaña.
- equipo complementario: adaptadores y ladrones de corriente, estación puertos USB, cargadores (USB) y cables de conexión. Este tipo de equipo es portátil y necesario para poder conectar varios dispositivos tanto para la carga de energía como para la elicitación y el acceso a los datos. Si fuera necesario, un equipo de carga solar es también recomendable. Esta última lista está abierta para cubrir cualquier condición de campaña; es decir, que, además de estos recursos, otros

Departamento de Language & Cognition del Max Planck Institute for Psycholinguists de Nijmegen, Países Bajos, donde se puede acceder a los manuales y materiales de elicitación de este grupo desde 1992 a 2010, para comprobar cómo se desarrollan este tipo de protocolos. El sitio web, Herramientas para la lingüística de campo [<https://www.eva.mpg.de/lingua/tools-at-lingboard/tools.php>] del Departamento de Lingüística del Max Planck Institute for Evolutionary Anthropology de Leipzig es también un buen recurso para encontrar materiales de elicitación e información fundamental para desarrollar el trabajo de campo.

3.1.2. Fase 2: el momento clave o durante la investigación *in situ*

Esta fase corresponde al momento álgido del trabajo de campo: cuando se llega al sitio de investigación y uno se prepara para iniciar la campaña de recogida de datos. Quizás, antes de explicar los pasos a seguir durante la campaña, sea necesario mencionar ciertos aspectos (consejos, indicaciones), menos científicos y más emocionales, de los que el investigador ha de ser consciente: (i) se ha de estar preparado para cualquier eventualidad, (ii) la paciencia y la positividad son fundamentales, (iii) los planes a veces no salen como uno ha planeado cuidadosamente, pero (iv) lo inesperado es a veces mejor que lo esperado.

Una vez armados de estas indicaciones generales, es momento para desgranar algunos elementos básicos del momento de recogida de datos. Algunos de estos elementos y su duración van a depender del tipo de trabajo de campo que se vaya a realizar. Por ejemplo, si se va a estar poco tiempo y se conoce de antemano a la comunidad lingüística, los preliminares de la sesión de trabajo, siempre obligatorios, pueden ejecutarse más rápido que si se está durante una larga temporada en dicha comunidad o es la primera vez que se colabora con esos informantes. En estos últimos casos, será necesario más tiempo de adaptación y aclimatación por parte de sendas partes: el investigador y la comunidad lingüística. En cualquier caso, y suponiendo que ambas partes están listas, las sesiones de trabajo podrían describirse de la siguiente manera.

3.1.2.1. Preliminares básicos

Una vez en el sitio de investigación ha de buscarse el lugar o el **espacio de trabajo** que se va a usar para las sesiones. En la Figura 8 hay varias instantáneas de diferentes espacios de trabajo en cinco zonas del mundo diferentes clasificadas en dos grupos según la perspectiva de la que suscribe este artículo: (i) zonas lejanas, como Ghana (a), Papúa Nueva Guinea (c), Brasil (e), México (f) y (ii) zonas cercanas, como Bizkaia (b) y Navarra (d).



Figura 8. Espacios de trabajo por el mundo

Desde luego va a depender mucho de dónde se esté o de la disponibilidad, pero, como norma general, es importante encontrar un lugar donde se puedan disponer adecuadamente el equipo de grabación (sonido, luz, encuadre...) y los utensilios que se necesiten para las tareas y, especialmente, donde el informante se encuentre a gusto y, de alguna manera, casi llegue a olvidarse de que está siendo “entrevistado” (si se está haciendo alguna tarea en concreto) u “observado” (si se está recogiendo material naturalístico, es decir, al natural, sin, apenas o ninguna, indicación o tarea específica). Asimismo, es muy conveniente mantener lo que se suele llamar un **diario de campo** que sirve para anotar todos los detalles de las sesiones de trabajo (fecha, hora, lugar, informante, tarea...) y cualquier incidencia o dato importante que surja durante la sesión. Estos diarios de campo suelen dar lugar, junto a las anotaciones y los análisis preliminares, a las conocidas como “notas de campo”.

Una vez se tiene todo dispuesto para iniciar la sesión de trabajo, se suele comenzar con un **preámbulo** para preparar al informante. Estos momentos sirven para al menos dos objetivos: (i) hacer que la situación se convierta en lo más natural posible (es decir, “romper el hielo” y que el informante se relaje, se olvide de la grabación) y (ii) llevar a cabo las explicaciones pertinentes. En este sentido, dependiendo del tiempo de campaña, especialmente si es corta, será el momento, también, de que quede plasmado en la grabación el consentimiento informado y la biografía del informante. Dependiendo del tipo de informante y de la sesión de elicitación, estos dos pasos se pueden hacer en diferentes momentos. Aunque a veces la parte previa a la ejecución de la tarea pueda parecer poco productiva, puesto que aún no se ha empezado a pasar las tareas para las que está programada la sesión, son muy útiles ya que, en muchos casos, junto con la información biográfica, pueden salir datos interesantes sobre la propia vida del informante, sobre su comunidad lingüística y, en muchas ocasiones, sobre diferentes ideologías y actitudes lingüísticas.

3.1.2.2. Durante la sesión de trabajo

La sesión de trabajo va a depender del tipo de tareas que se tengan previstas para ese día por lo que es difícil describir de manera específica qué hacer³. Por lo general, especialmente si la tarea no es manual, es decir, si consiste en describir diferentes materiales (imágenes, videos...) o responder a cuestionarios, la **duración** de la sesión incluyendo los preámbulos no debería de exceder más de una hora. Pese a que puede parecer poco tiempo, en realidad es mucho, teniendo en cuenta que los espacios de atención suelen oscilar alrededor de los quince minutos. Aunque todo depende de los estímulos y las tareas que se estén llevando a cabo, es recomendable intentar mantener el interés y la atención del informante en todo momento. Esto se puede hacer alternando la ejecución de diferentes tipos de tareas (si son cortas) y, sobre todo, acompañando al participante en todo momento para mostrarle que todo lo que está diciendo es interesante. En ocasiones cuando se repite una misma tarea con diferentes informantes es difícil para el investigador mantener la **concentración** puesto que ya ha sido testigo de la misma situación muchas veces. Esto no es admisible ya que puede afectar al informante de forma muy negativa. Una sesión de trabajo es siempre colaborativa por lo que, si una de las partes necesita descansar, es necesario parar. Por último, durante la sesión, el investigador, a veces, puede tomar notas en su diario de campo. Sin embargo, esta acción dependerá del tipo de informante, porque suele ser bastante disruptivo ver que alguien escribe mientras otro realiza una tarea.

3.1.2.3. Después de la sesión de trabajo

Aunque las sesiones de trabajo puedan ser agotadoras, es crucial procesar los datos recogidos en cuanto se termina. La palabra **procesamiento** de datos se usa de forma global para incluir desde la codificación y almacenamiento de las grabaciones, hasta la organización de las notas de campo y el comienzo de la transcripción de los datos. Aunque no siempre sea posible completar dichas tareas justo después de la sesión de trabajo, siempre es aconsejable adelantar parte del trabajo tras la sesión. Esto es importante no solo por la importancia que supone preservar el trabajo realizado, sino también por reflejar impresiones, detalles, ideas, anotaciones mentales, que han surgido durante la sesión, y que, por desgracia, permanecen durante muy poco tiempo, y, desde luego, que se olvidan si no se dejan registradas en las notas de campo.

En cualquier caso, lo que siempre se ha de hacer es la clasificación y el archivo de las grabaciones en diferentes soportes para ir confeccionando el **corpus** de datos. La **clasificación** consiste en la asignación del código a cada informante y a cada una de la(s) tarea(s) realizada(s). Como se mencionaba más arriba, el investigador habrá tomado nota de los detalles de la sesión en su diario de campo, pero ahora es el momento de ordenar las primeras notas y clasificar cada grabación. Esto es importante para poder localizar los datos de forma rápida y para después, en la producción, poder identificar el ejemplo que se vaya a usar dentro del corpus. Una vez que se han clasificado dichas grabaciones viene el **almacenaje** y el guardado de los archivos. Cualquier persona que trabaje con datos sabe que es importante guardarlos en diferentes dispositivos. Por esta razón, no se va a insistir en este aspecto. Tras este paso indispensable, lo siguiente sería comenzar con la transcripción de los datos de los archivos. Esta tarea es conveniente realizarla *in situ* puesto que puede darse el caso de tener que repetir, ampliar o consultar algún aspecto de la información recogida. En las campañas en las que se cuenta con un colaborador, tanto

un lingüista interno como un ayudante lingüístico, es crucial hacerlo en ese momento por razones obvias.

A la hora de hacer la **transcripción**, además de los detalles de identificación de cada informante y tarea, se pueden utilizar las diferentes herramientas y softwares que existen hoy en día para editar y transcribir archivos multimedia. La lista es amplia, pero quizás por ser más conocidos y utilizados, se pueden mencionar dos (véase sección 7 para su acceso en línea). Por un lado, el sistema de transcripción CHAT dentro del software CLAN que suele ser utilizado para repositorios como el *CHILDES* (véase MacWhinney 2000 y la sección 3.1.3.). Por otro lado, el software para anotación y transcripción multimodal ELAN (Lausberg y Sloetjes 2009) que es uno de los más frecuentes en la documentación lingüística por su fácil manejo y por las diversas opciones de alineado de texto oral, imagen y codificación que ofrece (esto es, cuenta con apartados específicos para glosado, traducción y cualquier otro tipo de información). En este punto, es importante que el investigador deje constancia del tipo de transcripción y transliteración que va a emplear, ya que muchas veces no está del todo establecida o estandarizada. Asimismo, para glosar los datos lingüísticos conviene utilizar sistemas que, más o menos, estén ya homogeneizados y aceptados por la comunidad científica como, por ejemplo, *The Leipzig Glossing Rules*, un conjunto de directrices para crear glosas de interlineado morfológico propuestas por el Max Planck Institute for Evolutionary Anthropology de Leipzig (Alemania).

3.1.3. Fase 3: la post-campaña o después de la recogida de los datos

Finalmente, la última fase comienza en el momento en el que se regresa del sitio de investigación y comprende al menos tres aspectos fundamentales: transcripción y análisis del corpus, almacenamiento permanente de los datos del corpus y la producción científica basada en dicha investigación. En la sección anterior, ya se han comentado algunos de los aspectos principales del primer aspecto y las razones para intentar llevar a cabo esta transcripción lo antes posible. Naturalmente la revisión de las transcripciones, la codificación y el análisis de los datos, especialmente esto último, son tareas que se extenderán en el tiempo, no solo porque son arduas y requieren una gran inversión de tiempo y esfuerzo, sino también porque, si están bien codificadas, se podrán explotar, de forma indefinida, desde diferentes ángulos y para multitud de objetivos de producción. Esta última sección, por tanto, se centrará exclusivamente en los otros dos aspectos.

En primer lugar, el **archivamiento** y el **almacenaje** de los datos (en crudo y transcritos) de forma consistente, regulada y permanente es crucial. Como se explicaba en la sección anterior, a cada informante y a cada sesión de trabajo se le asigna un código intransferible que servirá para identificar cada archivo (y después, para hacer referencia a esos datos y que sean fácilmente explotables y localizables). Estos archivos se han de guardar, por consiguiente, junto con toda la información psico-social correspondiente al informante (cuestionario biográfico) y al manual de campo.

Hoy en día, para almacenar este tipo de datos existen numerosos **repositorios** y plataformas de investigación de diferente tipología. Por ejemplo, entre las más conocidas (véase información en la sección 8), además de los archivos analógicos en museos y fundaciones (p. ej., *Smithsonian Institution*, *British Library*), se pueden citar: el repositorio *ELAR* del Programa de Documentación de Lenguas en Peligro de Extinción (véase la Figura 9), el repositorio *TLA* y *DOBES* del Max Planck Institute o el banco de

datos *TalkBank*, que acoge diferentes tipos de repositorios de interacción oral provenientes de diferentes metodologías incluida la del trabajo de campo. Así mismo, dentro del *TalkBank* e impulsado por Brian MacWhinney desde la Universidad Carnegie Mellon de Estados Unidos, también se encuentra la base de datos de interacciones infantiles *CHILDES* (MacWhinney 2000).



Figura 9. El repositorio ELAR. 550 lenguas en más de 70 países. Adaptado de <https://www.elararchive.org/>

Estas plataformas contribuyen no solamente a que los corpus queden custodiados por entidades públicas o privadas, sino para que, además, puedan ser utilizados por otros investigadores bajo demanda y con diferentes tipos de licencia. En este sentido, aunque la legislación sobre los derechos de propiedad intelectual en relación con las lenguas está aún en desarrollo –el concepto de quién es el dueño de una lengua es terriblemente complejo– y depende de cada país/comunidad, existen licencias como la *Creative Commons* (CC) o *Traditional Knowledge* (TK) que se utilizan para proteger el uso de este tipo de datos e información. En algunos casos, se pueden también seguir y adaptar las directrices de guías, declaraciones o códigos de ética ya establecidos para ciertas comunidades específicas donde se establecen bases generales tanto de protección como de uso y disfrute de los datos recogidos en campañas. Por ejemplo, la Declaración de las Naciones Unidas sobre los Derechos de los Pueblos Indígenas (ONU 2007), la Guía para la Conducta Ética de la Asociación de Estudios Africanos (ASA <https://africanstudies.org/about-the-asa/asa-ethical-conduct-guidelines/>), la colección de los Códigos de ética y profesionalidad de la Asociación Americana de Antropología (AAA: <https://archaeologicaethics.org/code-of-ethics/>) o la Guía para la investigación ética en los Estudios Indígenas del Instituto Australiano del Estudios Aborígenes y de las Islas del Estrecho de Torres (AIATSIS, <https://aiatsis.gov.au/research/ethical-research>),

entre otras. En cualquier caso, es responsabilidad del investigador tener claras estas restricciones y, sobre todo, respetarlas.

En segundo lugar, hay que mencionar la **producción** derivada del trabajo de campo. En este sentido, se pueden distinguir al menos tres tipos de producción que, en teoría, deberían siempre cubrirse de una manera u otra: (i) producción científica para foros especializados, (ii) divulgación científica para la sociedad del conocimiento en general, y (iii) transferencia de resultados para la comunidad lingüística en la que se ha trabajado. Los dos primeros casos están claros y suelen ser las vías centrales de explotación por parte del investigador. Sin embargo, el tercer caso suele desarrollarse en menor medida. Por el concepto de transferencia aquí se describe cualquier acción que revierta en el beneficio y el empoderamiento de la comunidad lingüística, desde la creación de materiales didácticos para la revitalización de la lengua hasta la exposición en abierto de los materiales recogidos. Cada una de estas vías de explotación va a depender de las condiciones específicas de la investigación llevada a cabo en el trabajo de campo. No obstante, en cualquiera de los tres casos, especialmente en los dos primeros, es imprescindible que el investigador reconozca y agradezca la colaboración con la comunidad lingüística de forma explícita en las publicaciones (o en casos en los que se ha tenido un colaborador experto incluirlo como co-autor) y que no se olvide de garantizar el acceso a estas publicaciones a dicha comunidad.

En definitiva, es conveniente concluir esta sección reproduciendo en la Tabla 1 los cinco principios éticos del trabajo de campo del investigador que resume Dwyer (2006: 38-40) y que deberían de estar siempre presentes en todas las fases del trabajo de campo:

Principio 1. No hagas daño (aunque sea sin querer o sin intención)
Principio 2. Reciprocidad y equidad (la relación de investigación ha de ser consensuada, continuamente negociada y respetuosa)
Principio 3. Haz algún bien (tanto a la comunidad como a la ciencia)
Principio 4. Obtén el consentimiento informado antes de empezar la investigación
Principio 5. Almacena y difunde tus datos y tus resultados

Tabla 1. Cinco principios de investigador en el trabajo de campo según Dwyer (2006)

4. Conclusiones: ¿Qué aporta a la lingüística el trabajo de campo?

En este artículo se ha presentado el concepto de trabajo de campo como una metodología lingüística necesaria para el avance de la comprensión del funcionamiento de las lenguas y del procesamiento del lenguaje. Las lenguas representan, crean y recrean la cosmovisión de sus hablantes. Son incardinaciones del saber humano que están enraizadas en el contexto sociocultural de sus respectivas comunidades lingüísticas, por lo que para poder descifrar cómo funciona el lenguaje, es imprescindible la aplicación de metodologías lingüísticas en las que el investigador tenga contacto de primera mano con el fenómeno que está investigando en su plenitud: el lenguaje y sus principales protagonistas, los hablantes.

Se ha propuesto definir el trabajo de campo de una manera general y amplia como: la recogida sistemática de información sobre un determinado fenómeno *in situ*, en su entorno natural. Es decir, se considera trabajo de campo cualquier tipo de investigación en la que el investigador se desplace al entorno natural (cercano o lejano) de la comunidad

lingüística con la que va a colaborar. Queda excluida, por tanto, de esta definición el trabajo empírico que se realiza con informantes en un laboratorio o por medios telemáticos –aunque desde el inicio de la pandemia del Covid-19 se han tenido que adaptar las campañas de documentación a la situación de emergencia (véase Orjuela et al. 2021)– puesto que no es posible cumplir algunas de las premisas básicas establecidas. Entre las principales ventajas que ofrece el trabajo de campo al estudio del lenguaje y las lenguas está precisamente el contacto directo con la comunidad lingüística. Esto supone el acceso no solamente a materiales de primera mano, reales y contrastados con sus usuarios, sino también contextualizados, es decir, recogidos en su propio hábitat, lo que a veces es crucial para poder interpretarlos de forma adecuada. Un ejemplo para ilustrar la importancia de este tipo de metodología se puede ver en la Figura 10.

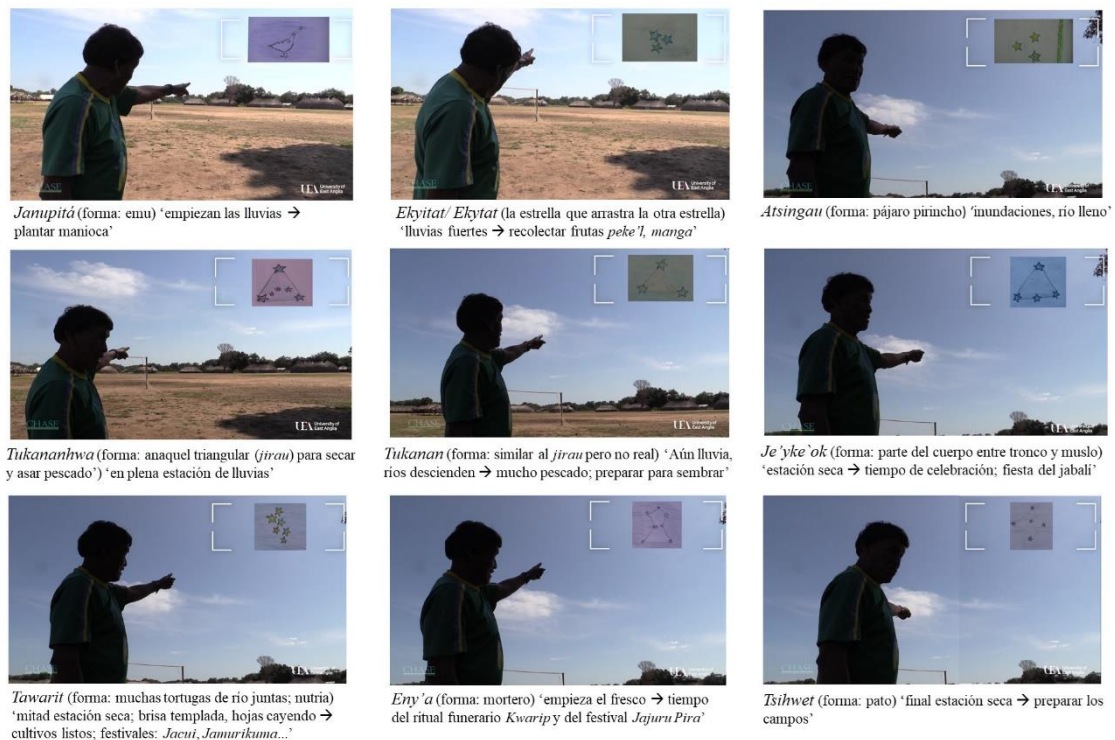


Figura 10. Las constelaciones y los eventos en kamaiurá. Dibujos de Wary Kamaiurá Sabino. Adaptado de Silva-Sinha (2018: 154-155; documental *Time in Culture*)

Las imágenes de la Figura 10 representan las diferentes constelaciones de estrellas y cómo se sitúan y se interpretan en el cielo de los kamaiurá, vistas desde el pueblo de Ipawu en el Alto Xingú de la provincia del Mato Grosso en Brasil. Esta comunidad, como describe Vera da Silva-Sinha (2018, 2019), organiza y conceptualiza el tiempo a partir de eventos, es decir, a partir de sucesos que ocurren. El tiempo no se puede “medir” en horas o meses, como se hace en occidente con artefactos culturales como los relojes o los calendarios, sino que se organiza alrededor de intervalos de sucesos. Por ejemplo, se sirven de las diferentes etapas de madurez de una persona para determinar la “edad” de ese individuo en vez de la fecha de cumpleaños, utilizan las fases de la luna para marcar el ciclo menstrual de la mujer (véase la Figura 8.c.) o, como en la Figura 10, miran la posición y el aspecto de las estrellas y sus constelaciones para establecer la temporalización de las cosechas y las celebraciones. Aunque, por supuesto, se puede comprender el sistema temporal de los kamaiurá desde fuera de este entorno⁴, sería tremendamente complejo

unir cada constelación, su nombre metonímico basado en la forma de los elementos de los que toman su nombre, su significado y su impacto en la vida de los kamaiurá sin estar *in situ* y sin ver cómo se explica oral y gestualmente la conceptualización del tiempo.

Ahora bien, en el trabajo de campo es crucial no olvidar que este “acceso VIP” a las lenguas conlleva también (i) una serie de obligaciones con respecto a la comunidad lingüística en la que se trabaja (protocolos, respeto y empatía lingüística, documentación y materiales...), (ii) unos requisitos mínimos para el diseño de una campaña de trabajo adecuada desde sus pasos iniciales (contacto, diseño de tareas/materiales...) hasta el procesamiento de los datos (análisis, producción, archivo, conservación, disfrute), pasando por la recogida de los mismos (grabación, transcripción) y (iii) un esfuerzo extra por la inversión de tiempo y fondos necesarios para llevar a cabo este tipo de metodología.

En estas páginas, solo se han ofrecido algunas pinceladas de las bases del trabajo de campo, pero actualmente existe una abundante literatura especializada donde encontrar información más detallada. Por ejemplo, hay disponibles referencias específicas para ampliar conocimientos (i) sobre cómo llevar a cabo la documentación y la elaboración de gramáticas descriptivas de lenguas, uno de los objetivos clásicos en los que se aplica el trabajo de campo y que se suele considerar un campo de especialización propio (Vaux y Cooper 1998; Gippert et al. 2006; Vaux et al. 2007; Bowerman 2008; Chelliah y de Reuse 2011), (ii) sobre aspectos más técnicos y/o menos explorados del trabajo de campo (Thieberger 2012), o (iii) sobre la aplicación de esta metodología en comunidades lingüísticas remotas (Crowley 2007; Meakins et al. 2018), etc. El lector también puede consultar algunas referencias clásicas, como Samarin (1967), Bouquiaux y Thomas (1972) y Kibrik (1977), entre otros, y que son imprescindibles para iniciarse y conocer los comienzos de esta área. O incluso pueden consultarse publicaciones que se mueven entre la exposición científica de resultados del trabajo de campo y la narración por el viaje interior e íntimo del investigador de campo (Newman y Ratliff 2001; Harrison 2007; Evans 2008, 2011; Everett 2008; Sarvasy y Forker 2018, entre otros).

Sea como fuere, el trabajo de campo, en todas sus dimensiones (remoto, urbano, explorador, efímero...) es una metodología lingüística que no solo es fundamental para el avance del conocimiento sobre el lenguaje humano, sino que es, sin lugar a dudas, imprescindible para la formación de todo aquel que quiera ser **Lingüista**.

5. Referencias

- Aliaga Jiménez, José Luis; Giralt Latorre, Javier; Moret Oliver, María Teresa. 2016. Tres hitos en la historiografía de las lenguas minoritarias de Aragón. En M.^a C. Horno Chéliz, I. Ibarretxe-Antuñano y J. L. Mendivil-Giró, eds. *Panorama actual de la ciencia del lenguaje. Primer sexenio de Zaragoza Lingüística*. Zaragoza: PUZ, pp. 403-427.
- Amorós Negre, Carla. 2014. *Las lenguas en la sociedad*. Madrid: Síntesis.
- Aurrekoetxea, Gotzon. 2022. Métodos de la dialectología cuantitativa. *Estudios de Lingüística del Español* 45: 43-69.
- Austin, Peter K.; Sallabank, Julia, eds. 2011. *The Cambridge handbook of endangered languages*. Cambridge: Cambridge University Press.

- Barwick, Linda. 2012. Including music and the temporal arts in language documentation. En N. Thieberger, ed. *The Oxford handbook of linguistic fieldwork*. Oxford: Oxford University Press, pp. 166-179.
- Bouquiaux, Luc; Thomas, Jacqueline M. C.; Robert, James. 1972. *Enquête et description des langues à tradition orale*. Paris: Société d'Études Linguistiques et Anthropologiques de France.
- Bowern, Claire. 2008. *Linguistic fieldwork: A practical guide*. London: Palgrave Macmillan.
- Browne, Wayles. 2020. Bosnian-Croatian-Montenegrin-Serbian language. *Encyclopedia Britannica*. Disponible en: <https://www.britannica.com/topic/Bosnian-Croatian-Montenegrin-Serbian-language>.
- Cámara-Leret, Rodrigo; Bascompte, Jordi. 2021. Language extinction triggers the loss of unique medicinal knowledge. *Proceedings of the National Academy of Sciences* 118 (24) e2103683118. DOI: 10.1073/pnas.2103683118
- Chelliah, Shobhana L.; Reuse, Willem J. de. 2011. *Handbook of descriptive linguistic fieldwork*. Dordrecht: Springer Science + Business Media.
- Crowley, Terry. 2007. *Field linguistics: A beginner's guide*. Oxford: Oxford University Press.
- Dingemans, Mark; Verhoef, Tessa; Roberts, Seán G. 2014. The role of iconicity in the cultural evolution of communicative signals. En B. de Boer y T. Verhoef, eds. *Proceedings of EvoLang X workshop on signals, speech and signs*. Vienna: Department of English and American Studies, pp. 11-15.
- Dunn, Michael; Majid, Asifa; Jordan, Fiona. 2010. *Biographical questionnaire for the EoSS project. Incorporates the Edinburgh handedness inventory (Oldfield, 1971)*. Nijmegen: Max Planck Institute for Psycholinguistics.
- Dwyer, Arienne M. 2006. Ethics and practicalities of cooperative fieldwork and analysis. En J. Gippert, N. P. Himmelmann y U. Mosel, eds. *Essentials of language documentation*. Berlin: Mouton de Gruyter, pp. 31-66.
- Evans, Nicholas. 2008. Review of essentials of language documentation. *Language Documentation & Conservation* 2.2: 340-350.
- Evans, Nicholas. 2011. *Dying words: Endangered languages and what they have to tell us*. Malden: John Wiley & Sons.
- Epps, Patience; Webster, Anthony K.; Woodbury, Anthony C. 2017. A holistic humanities of speaking: Franz Boas and the continuing centrality of texts. *International Journal of American Linguistics* 83.1: 41-78.
- Everett, Daniel L. 2008. *Don't sleep, there are snakes: Life and language in the Amazonian jungle*. New York: Pantheon Books.
- Gippert, Jost; Himmelmann, Nikolaus P.; Mosel, Ulrike, eds. 2006. *Essentials of language documentation*. Berlin: Mouton de Gruyter.
- Hachimi, Atiqa. 2015. "Good Arabic, Bad Arabic": Mapping language ideologies in the Arabic-speaking world. *Zeitschrift für Arabische Linguistik* 61: 36-71.
- Harrison, K. David. 2007. *When languages die. The extinction of the world's languages and the erosion of human knowledge*. Oxford: Oxford University Press.
- Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36: 161-195.
- Hyman, Larry M. 2001. Fieldwork as a state of mind. En P. Newman y M. Ratliff, eds. *Linguistic fieldwork*. Cambridge: Cambridge University Press, pp. 15-33.

- Ibarretxe-Antuñano, Iraide. 2020. Empatía lingüística. *Revista Archiletras* 9: 70-72.
<https://www.archiletras.com/firma/empatia-linguistica/>
- Ibarretxe-Antuñano, Iraide. 2021a. Érase una vez... la estandarización de la lengua. *Revista Archiletras* 13: 58-59.
- Ibarretxe-Antuñano, Iraide. 2021b. De los hablantes a las lenguas pasando (o no) por la estandarización: un camino de ida y vuelta. *Zaragoza Lingüística a la Carta*.
<https://sites.google.com/view/zl-a-la-carta/zl-a-la-carta/más-allá-de-zaragoza-lingüística/de-los-hablantes-a-las-lenguas-pasando-o-no-por-al-estandarización?authuser=0>.
- Ibarretxe-Antuñano, Iraide; Valenzuela Manzanares, Javier. 2021. *Lenguaje y cognición*. Madrid: Síntesis.
- Kibrik, A. E. 1977. *The methodology of field investigations in Linguistics. (Setting up the problem)*. The Hague: Mouton.
- Lausberg, Hedda; Sloetjes, Han. 2009. Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments & Computers* 41.3: 841-849. DOI: 10.3758/BRM.41.3.841
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- Meakins, Felicity; Green, Jennifer; Turpin, Myfany. 2018. *Understanding linguistic fieldwork*. London: Routledge.
- Moreno Cabrera, Juan Carlos. 1997. *Introducción a la lingüística. Enfoque tipológico y universalista*. Madrid: Síntesis.
- Newman, Paul; Ratliff, Martha, eds. 2001. *Linguistic fieldwork*. Cambridge: Cambridge University Press.
- Ramsey, S. Robert. 1987. *The languages of China*. Princeton, NJ: Princeton University Press.
- Thieberger, Nick, ed. 2012. *The Oxford handbook of linguistic fieldwork*. Oxford: Oxford University Press.
- Sakel, Jeannette; Everett, Daniel L. 2012. *Linguistic fieldwork: A student guide*. Cambridge: Cambridge University Press.
- Samarin, William J. 1967. *Field linguistics. A guide to linguistic field work*. New York: Holt, Rinehart and Winston.
- Sarvasy, Hannah; Forker, Diana. 2018. *Word hunters: Field linguists on fieldwork*. Amsterdam: John Benjamins.
- Seyfeddinipur, Mandana. 2012. Reasons for documenting gestures and suggestions for how to go about it. En N. Thieberger, eds. *The Oxford handbook of linguistic fieldwork*. Oxford: Oxford University Press, pp. 147-165.
- Silva-Sinha, Vera. 2018. *Linguistic and cultural conceptualisations of Time in Huni Kuĩ, Awetý and Kamaiurá indigenous communities of Brazil*. Tesis doctoral inedita. University of East Anglia.
- Silva-Sinha, Vera. 2019. Event-based time in three indigenous Amazonian and Xinguan cultures and languages. *Frontiers in Psychology*. DOI: 10.3389/fpsyg.2019.00454
- UNESCO. 2003. Vitalidad y peligro de desaparición de las lenguas: grupo especial de expertos sobre las lenguas en peligro convocado por la UNESCO. Disponible en: http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CLT/pdf/LVE_Spanish_EDITED%20FOR%20PUBLICATION.pdf

- ONU. 2007. Declaración de las Naciones Unidas sobre los Derechos de los pueblos indígenas. Disponible en:
https://www.un.org/development/desa/indigenouspeoples/wp-content/uploads/sites/19/2018/11/UNDRIP_S_web.pdf
- Orjuela, Lorena; Ortiz Álvarez, Rebeca; Galván, Johanna. 2021. Documentar una lengua en tiempos de SARS-CoV-2. *Visitas al Patio* 15.2: 212-232.
- Vaux, Bert; Cooper, Justin. 1998. *Introduction to linguistic field methods*. Munich: Lincom Europa.
- Vaux, Bert; Cooper, Justin; Tucker, Emily. 2007. *Linguistic field methods*. Eugene, OR: Wipf and Stock.
- Vicente, Ángeles. 2011. La diversidad de la lengua árabe como lengua de comunicación. *Miscelánea de Estudios Árabes y Hebraicos, Sección Árabe e Islam* 60: 353-370.

6. Sitios web con recursos y/o información sobre el Trabajo de Campo y sobre la diversidad lingüística

Información sobre Trabajo de Campo (recursos, formación, financiación). Endangered Languages Documentation Programme. Disponible en: <https://www.eldp.net/>

Información sobre Trabajo de Campo. Stanford University. Disponible en: <http://www.stanford.edu/dept/linguistics/cgi-bin/fieldwork/info.php>

Información sobre la preparación del Trabajo de Campo. University of Toronto. Disponible en: <http://projects.chass.utoronto.ca/lingfieldwork/>

Herramientas tipológicas para el Trabajo de Campo. Max Planck Institute, Leipzig. Disponible en: <http://www.eva.mpg.de/lingua/tools-at-lingboard/tools.php>

Manuales y estímulos para el trabajo de campo (instrucciones, materiales, etc.). Max Planck Institute, Nijmegen. Disponible en: <http://fieldmanuals.mpi.nl/>

Información sobre las lenguas del mundo: Glottolog 4.5 (Gratuito). Disponible en: <https://glottolog.org/>

Información sobre las lenguas del mundo: Ethnologue (Bajo suscripción). Disponible en: <http://www.ethnologue.com>

7. Software

ELAN. Software para la anotación de grabaciones en audio y video. Max Planck Institute. (Lausberg y Sloetjes 2009). <https://archive.mpi.nl/tla/elan>

CLAN. Software para el análisis computerizado del lenguaje a través de CHAT (MacWhinney 2000). <https://dali.talkbank.org/clan/>

8. Archivos

DOBES. Documentación de Lenguas Amenazadas. Max Planck Institute, Nijmegen.
<https://dobes.mpi.nl/>

ELAR. Endangered Language Archive. Endangered Language Documentation Programme. <https://www.elararchive.org/>

TalkBank. Brian MacWhinney. Carnegie Mellon University, EUA. <https://talkbank.org/>

TLA. The Language Archive. Max Planck Institute, Nijmegen <https://archive.mpi.nl/tla/>

* Este trabajo se encuadra dentro de: los proyectos CONESSO (FFI2017-82460-P) y MultiMetAR (DGA-LMP143_21) y los grupos de investigación Psylex (UZ-DGA H11_17R) y ICON (Campus Íberus). Gracias infinitas a las editoras, a los revisores externos y también a los colegas que me han facilitado las pruebas fotográficas de su trabajo de campo (y a sus (y mis) respectivas comunidades lingüísticas): Juergen Bohnemeyer (lengua: maya yucateco; Yaxley, México); Mark Dingemanse (lengua: siwu; Akpafu-Mempeasem, Ghana); Günter Senft (lengua: kilivila; Tauwema, Islas Trobiand, Papúa Nueva Guinea); Vera da Silva Sinha (lenguas: awetý, kamaiurá; reserva Xingú, Brasil); Julia Guenbelzu (lengua: español, Zaragoza, Aragón); María Louro-Mendiguren (lengua: euskara, Ondarroa, País Vasco) y una servidora (lengua: euskara-castellano; Ikastola Astileku, Portugalete, País Vasco; lengua: euskara, Orbaizeta, Navarra; lengua: aragonés, Ansó, Aragón).

¹ En muchos de estos casos, estas denominaciones vienen determinadas por agendas geopolíticas, económicas, religiosas, etc. de gobiernos, entidades o grupos con algún tipo de influencia o poder en distintos momentos históricos, no necesariamente actuales. Al ser estos aspectos extralingüísticos quedan fuera del alcance de esta sección; no obstante, es crucial no obviarlos puesto que, en el trabajo de campo, al estar (o querer estar) el investigador en contacto con una comunidad lingüística es necesario conocer a fondo todo el contexto, actual y pasado, en el que se desarrolla la lengua de su investigación (véase sección 2.2.). Casos ilustrativos de este tipo de situaciones hay muchos en el mundo. Por ejemplo, la elección de apelativos para nombrar a la lengua/dialecto para separarse o unirse a un grupo de variedades por motivos de identidad política (p. ej., el caso del catalán vs. el valenciano; o del xapurreo y del acrónimo LAPAO (Lengua Aragonesa del Área Oriental) para evitar la etiqueta catalán de Aragón (véase Aliaga Jiménez et al. 2016)) o por motivos de identidad religiosa (p. ej., árabe (clásico) vs. variedades, Hachimi 2015; Vicente 2011). Aunque, quizás uno de los casos más ilustrativos, porque aúna todos estos factores (geopolíticos, identidad, religión...) sea el del serbocroata que ha transitado por diferentes fases: desde sus inicios históricos en el eslavo eclesiástico antiguo (s. X), pasando por su acuerdo de estandarización (s. XIX) y su abandono tras la desaparición de Yugoslavia (s. XX), hasta los debates actuales sobre su verdadero o falso carácter de lengua policéntrica en relación a la lengua BCMS [BosnioCroataMontenegrinoSerbio] (Browne 2020).

² En este artículo se da por hecho que el investigador tiene claro el fenómeno lingüístico o la lengua a investigar, por eso no se dan más detalles de cómo escogerlos. No obstante, en los manuales más completos sobre trabajo de campo, como en Chelliah y de Reuse (2011: caps. 4 y 5) o en Meakins et al. (2018: sección 1.3., 2.1.), se incluye información específica al respecto.

³ En este artículo no hay espacio para describir las diferentes tareas que se pueden realizar sobre las diversas áreas de los estudios lingüísticos. El lector puede consultar manuales, como Chelliah y de Reuse (2011), Bovern (2015) o Meakins et al. (2018), donde se dedican capítulos individuales al trabajo de campo en temas fonético-fonológicos, morfosintácticos, semánticos, pragmáticos, sociolingüísticos, etc. Así mismo, manuales como Sakel y Everett (2012) y los recursos que se listan en la sección 6 (especialmente los del Max Planck Institute) pueden ser de utilidad para conocer diferentes tipos de metodología cuantitativa y cualitativa. Para una descripción de áreas menos exploradas, el lector puede consultar el manual editado por Thieberger (2012).

⁴ Además de las referencias de Silva-Sinha (2018, 2019), su documental *Time in culture* es una exquisita pieza para comprender este sistema temporal. Este documental está accesible en: <https://vimeo.com/261572557>.

Las tareas conductuales en la investigación sobre el procesamiento del lenguaje

José Manuel Igoa
Universidad Autónoma de Madrid
josemanuel.igoa@uam.es

Resumen

En este artículo se exponen las principales tareas conductuales empleadas en la investigación sobre el procesamiento del lenguaje en las áreas de comprensión y producción. El artículo tiene dos objetivos principales. El primero de ellos es exponer las razones por las que la investigación psicolingüística puede aportar datos de interés para la Lingüística teórica, sin dejar de reconocer algunas dificultades que complican esta contribución. Se arguye que la metodología y, en particular, las tareas o los procedimientos de recogida de datos pueden ayudar a paliar estos problemas. El segundo objetivo del artículo es describir un conjunto de tareas caracterizadas como conductuales que se pueden agrupar bajo diversos criterios de clasificación. Se presentan de forma sucesiva las tareas conductuales utilizadas en la investigación sobre los procesos de comprensión y las empleadas en la investigación de los procesos de producción del lenguaje, examinando sus ventajas e inconvenientes. Un aspecto importante en esta exposición es la noción de ‘paradigma experimental’, de la que se muestran dos destacados ejemplos: los paradigmas de *priming* y de atención dividida (o tarea dual), el primero de los cuales se emplea tanto en tareas de comprensión como de producción del lenguaje. El artículo concluye con unas consideraciones sobre las deficiencias de las tareas conductuales, si bien se defiende la idea de que estas deficiencias no son exclusivas de esta clase de tareas y se reivindica tanto la utilidad como la necesidad de las mismas en la investigación empírica del lenguaje.

Palabras clave: Psicolingüística, comprensión y producción del lenguaje, paradigmas experimentales, tareas sobre la marcha y a posteriori.

Abstract

This article presents the major behavioral tasks used in research on language processing in both comprehension and production. The article has two main goals. First, to justify the claim that psycholinguistic research can provide significant contributions to theoretical linguistics, despite a number of difficulties that complicate this endeavor. It is claimed that methodology, and particularly the choice of tasks and procedures for data collection, are useful tools to circumvent or overcome these difficulties. The second goal of the article is to describe a handful of tasks labelled behavioral by means of a few significant criteria selected for that purpose. A number of behavioral tasks commonly employed in research on language comprehension and production are described in turn, highlighting their main strengths and weaknesses. A major issue in this regard is the notion of ‘experimental paradigm’, of which two prominent examples are presented, namely the priming paradigm and the divided attention or dual-task paradigm, the former used both in comprehension and production tasks. The article ends with some remarks on the main shortcomings of behavioral tasks, and with the claim that these pitfalls are not

exclusive of this kind of tasks. Be that as it may, it is finally contended that behavioral tasks are useful and necessary in empirical research on language.

Keywords: Psycholinguistics, language comprehension and production, experimental paradigms, online and offline tasks.

1. Introducción

Del extenso conjunto de tareas y métodos de investigación que ofrece la Psicolingüística, las denominadas *tareas conductuales* constituyen un conjunto heterogéneo de procedimientos que se han agrupado bajo diversos criterios y se han definido con arreglo a consideraciones variopintas, hasta el punto de que solo los unen dos únicos hilos conductores: por un lado, la dependencia que presentan con respecto a la ejecución de respuestas voluntarias por parte de los participantes, y, por otro, el empleo del tiempo en iniciar o producir la respuesta y el grado de acierto en la tarea como variables dependientes o de respuesta más frecuentemente utilizadas. De hecho, a mi modo de ver, y a pesar de su extendido uso, la etiqueta “tareas conductuales” no refleja adecuadamente el referente al que pretende aludir, pues no todo lo que comprende son tareas propiamente dichas ni tampoco estas se distinguen de las que se emplean en otros métodos de investigación en Lingüística por el hecho de ser conductuales. Por lo tanto, lo que se esconde bajo este amplio paraguas son diversos procedimientos de recogida de datos que varían en el tipo y el grado de manipulación de los materiales y en las condiciones de observación empleados en la investigación, que abarcan desde la mera observación del comportamiento sin intrusión alguna hasta el minucioso control de los materiales empleados y de las condiciones bajo las cuales se registran las respuestas de los participantes (Garrod 2006; Grey y Tagarelli 2018).

El presente artículo está dividido en dos grandes apartados. En el primero, me propongo hacer una reflexión sobre los fundamentos de la investigación psicolingüística y los problemas y dificultades concomitantes a los mismos, con el objetivo de justificar la utilidad y la pertinencia de los procedimientos conductuales de investigación para ayudar a resolver problemas de interés para la Lingüística. Esto me llevará a examinar los supuestos en los que se asienta la investigación psicolingüística y los requisitos que debe cumplir, así como la estructura y los componentes que forman parte de las tareas que se emplean en este tipo de investigaciones. A partir de este marco de referencia global, en el segundo apartado se expondrán varias distinciones relevantes que nos servirán para clasificar los procedimientos conductuales, objeto central del artículo. De esta clasificación se seleccionarán y describirán varios casos representativos de las tareas conductuales que se utilizan en el ámbito de procesamiento del lenguaje en los dominios de la comprensión y la producción. El artículo concluirá con una valoración del alcance y limitaciones de las tareas examinadas y un comentario sobre la utilidad relativa de las mismas.

2. La investigación en Psicolingüística: fundamentos, problemas y distinciones relevantes

El propósito de este apartado es tratar de aclarar hasta qué punto la investigación empírica sobre el procesamiento del lenguaje puede aportar información relevante para la Lingüística teórica. Planteado el problema en estos términos, lo que interesa averiguar y, en su caso, justificar es en qué medida el estudio empírico de las operaciones que los hablantes oyentes realizamos en tiempo real al comprender y producir enunciados lingüísticos puede ayudar a entender la capacidad humana para el lenguaje, es decir, a descubrir el conocimiento implícito que tenemos del lenguaje en tanto que sistema de reglas y representaciones permanentes que subyacen y posibilitan el uso de la lengua. Sobre la naturaleza de este conocimiento pesan numerosos interrogantes¹ que compete responder a los lingüistas, pero la cuestión es si esta ingente tarea debe descansar en exclusiva sobre sus fatigados hombros o pueden compartirla con otros estudiosos de las ciencias del lenguaje, como los psicólogos o los científicos cognitivos en general. Si optamos por la tarea compartida, uno de los intereses que debe orientarla es el de contrastar la “realidad psicológica” de los constructos y entidades que postulan las teorías del lenguaje. No obstante, esto solo será posible si asumimos que existe alguna forma de correspondencia o interdependencia entre las reglas y representaciones del lenguaje en un plano abstracto y los procesos de uso en tiempo real, o, por usar la terminología popularizada por Chomsky, entre la competencia lingüística y la actuación (Lewis y Phillips 2015).

Dejando a un lado a los partidarios de la independencia entre competencia y actuación, podemos encontrar distintos grados de correspondencia entre ellas, desde una correspondencia parcial entre los constructos de la Lingüística y los procesos y representaciones psicolingüísticos, hasta un estricto paralelismo o correspondencia total entre ambos. Sea cual fuere la opción preferida, el estudio de los procesos psicolingüísticos se orienta hacia dos objetivos fundamentales: (i) conocer la naturaleza y el funcionamiento coordinado de estos procesos durante la ejecución y la comprensión de los actos de habla e (ii) indagar en el papel de los recursos de atención y memoria, entre otros, que hacen posible y a la vez restringen los procesos de actuación.

Sin embargo, la posible contribución de las teorías y los estudios sobre el procesamiento del lenguaje a la explicación del conocimiento del lenguaje tropieza con algunas dificultades que es preciso reconocer y resolver. La primera de estas dificultades reside en el hecho de que la Lingüística teórica pretende no solo desvelar la naturaleza del conocimiento del lenguaje, una tarea a la que las teorías psicolingüísticas y los datos empíricos pueden aportar información de interés, sino también analizar las propiedades lógicas del lenguaje. Ejemplos de estas son la sinonimia, la implicación, la contradicción o la gramaticalidad, propiedades que no dependen, en sentido estricto, de los juicios de los usuarios del lenguaje ni se pueden estudiar experimentalmente (Soames 1984). Poco o nada tiene que decir una ciencia empírica como la Psicolingüística sobre estas cuestiones.

Otro problema, de mayor envergadura que el anterior, es que el conocimiento del lenguaje no es estático, sino que cambia de manera sustancial a lo largo de los primeros años de la vida. Esto obliga a tener en consideración las relaciones que existen entre el procesamiento del lenguaje y los procesos de adquisición, una relación compleja que discurre en ambas direcciones. De hecho, algunas teorías influyentes en Psicolingüística sostienen que la adquisición del lenguaje depende directamente del uso (y, por tanto, del procesamiento) del lenguaje en situaciones de interacción comunicativa (Tomasello 2003) y, por extensión, que la gramática emerge, tanto en su evolución filogenética como

en su devenir histórico en las lenguas particulares, como resultado de la acumulación de cambios puntuales derivados del uso del lenguaje por los individuos pertenecientes a distintas comunidades de hablantes (Christiansen y Chater 2016).

Una tercera dificultad, bien conocida por los lingüistas, es que hay una asimetría entre el conocimiento y el procesamiento del lenguaje, dado que el procesamiento del lenguaje no es uno, puesto que se desdobra en tareas de comprensión y tareas de producción. Como consecuencia, existe la posibilidad de que pueda haber sistemas de conocimiento distintos implicados en cada una de estas tareas e incluso en función de las diferentes modalidades de externalización del lenguaje (oral, escrita, signada), al menos en lo que atañe a componentes específicos del conocimiento del lenguaje. Por ejemplo, diversos modelos de procesamiento derivados del estudio de las alteraciones neuropsicológicas del lenguaje, como las afasias o las dislexias adquiridas, postulan la existencia de hasta cuatro “lexicones” o almacenes de palabras, dos de entrada (para la comprensión oral y escrita) y dos de salida (para la producción oral y escrita) (Friedmann et al. 2013).

Por último, otro hecho bien conocido es que la competencia y la actuación lingüísticas difieren en que esta última está sujeta a interferencias y distorsiones, o también a decisiones estratégicas, que son fenómenos inherentes a la externalización del lenguaje y, por tanto, no afectan a la competencia. Este es el caso de la ambigüedad lingüística, que acontece cuando una determinada palabra o cadena de palabras se puede interpretar de varias formas distintas. La ambigüedad se puede dar en diferentes componentes o niveles de procesamiento del lenguaje, como el léxico –cuando una forma léxica se asocia con diversos significados–, la sintaxis –cuando a una cadena de palabras le corresponde más de una estructura posible– o la pragmática –cuando un enunciado lingüístico puede expresar diferentes intenciones comunicativas del hablante. El hecho mismo de que la información del contexto que rodea el elemento ambiguo se utilice con rapidez y eficacia para deshacer la ambigüedad es buena prueba de que se trata de un problema, generalmente transitorio, de la actuación, y no un fallo en la competencia (Meseguer y Carreiras 1999; Rodd et al. 2002; Wasow et al. 2005; Macagno y Bigi 2018). Otros casos bien conocidos de problemas atribuibles a los procesos de actuación son los errores perceptivos o de comprensión (como los que ocurren en las llamadas “oraciones de vía muerta”² –Fodor e Inoue 1994), los *lapsus linguae* o errores en la producción del habla (Del Viso 2002), la evidencia de que a menudo procesamos los enunciados de manera incompleta o superficial (Ferreira y Patson 2007) o la imposibilidad de procesar ciertas estructuras que desbordan nuestra capacidad de memoria de trabajo (como las oraciones con múltiple inclusión central³ –Fodor et al. 2017). Estas peculiaridades de la actuación pueden contaminar los datos empíricos emanados de la investigación si lo que se pretende con ellos es probar la realidad psicológica de las unidades, mecanismos, reglas o representaciones que postulan las teorías lingüísticas. El problema es que no siempre es posible distinguir los efectos genuinamente derivados del ejercicio de la competencia de aquellos que obedecen a estrategias o a errores atribuibles primordial o exclusivamente a la actuación.

Un medio para minimizar los problemas descritos reside en la metodología. El cometido de la metodología es seleccionar cuidadosamente los materiales que se van a emplear en la investigación, manipular de forma adecuada las variables que hipotéticamente intervienen en el proceso que se pretende examinar y controlar o suprimir otros factores que puedan desvirtuar las conexiones causales o correlacionales que existen entre los factores relevantes. En lo que atañe a los métodos de investigación en Psicolingüística, la

elección de la tarea que los participantes han de realizar es un elemento crucial de la metodología, dado que es el instrumento que permite reproducir la situación natural en la que se manifiesta el comportamiento objeto de estudio y que asegura la manipulación de las variables intervinientes y el control o la supresión de las contaminantes.

Las tareas comúnmente utilizadas en la investigación psicolingüística están diseñadas para obtener diversos tipos de medida de las respuestas de los participantes. La principal de estas medidas es el tiempo, ya sea en forma de latencia (o tiempo de reacción ante un estímulo) o de duración de la respuesta (Garrod 2006; Grey y Tagarelli 2018). Las medidas de tiempo se interpretan como un indicador de la complejidad o la dificultad de la tarea, bajo el supuesto de que existe una relación lineal entre el tiempo invertido en iniciar la respuesta y la complejidad de los procesos que subyacen a su preparación y ejecución. Sin embargo, hay abundante evidencia de que esta relación no es lineal (o dicho en términos técnicos, no se ajusta a una función matemática monótonamente creciente). Además, esta relación está mediada por un factor estratégico, el denominado “compromiso entre la rapidez y la precisión”, según el cual en las tareas que tienen un criterio de corrección, cada participante puede optar por anteponer la rapidez en la respuesta en perjuicio de la precisión o, al contrario, asegurar un mayor nivel de aciertos a costa de una mayor lentitud en la ejecución de la respuesta (Wickelgren 1977; Heitz 2014). Este factor estratégico debe ser tenido en cuenta y corregido para interpretar adecuadamente las medidas de tiempo. Además del tiempo y la precisión, hay otras medidas cuantitativas como la frecuencia o la proporción con la que se dan diferentes tipos de respuestas, previamente categorizadas, en el transcurso de la tarea. Por último, también se pueden analizar aspectos cualitativos de las respuestas de los participantes, tales como la topografía o configuración física, en el caso de respuestas motoras o en registros de la actividad neurológica, o la categorización semántica de respuestas verbales en tareas de elección múltiple que requieren argumentación o justificación.

A la hora de clasificar las tareas conductuales que se van a presentar en el apartado 3, es preciso prestar atención a una serie de distinciones que obedecen a diversos criterios. En primer lugar, podemos distinguir entre (i) tareas metalingüísticas, que requieren juicios conscientes y manifiestos acerca de determinadas propiedades de los materiales lingüísticos que se presentan en la tarea, como la gramaticalidad de los enunciados o su plausibilidad, y (ii) tareas orientadas a evaluar aspectos del procesamiento, que aunque puedan requerir juicios metalingüísticos, utilizan el tiempo de respuesta como indicador de la dificultad de la tarea. Por ejemplo, en la investigación sobre el reconocimiento y la comprensión de palabras, se suelen contrastar los resultados de dos tareas diferentes: por un lado, la tarea de denominación de palabras (o pseudopalabras), consistente en leer en voz alta una secuencia de letras que corresponde o no a una pieza léxica del idioma; y, por otro lado, la tarea de decisión léxica, que consiste en juzgar si la secuencia de letras que se presenta constituye o no una palabra⁴ (Balota y Chumbley 1984; Katz et al. 2011). A efectos de esta distinción, la segunda es una tarea metalingüística, mientras que la primera no lo es. Sin embargo, eso no impide que se puedan comparar los tiempos de respuesta de ambas, al ser la de denominación una tarea más superficial, sensible principalmente a la frecuencia de uso de las sílabas o secuencias de grafemas y a la frecuencia de la palabra, en contraste con la tarea de decisión léxica, que no solo es sensible a estos factores de frecuencia, sino también a las propiedades semánticas de la palabra.

La segunda distinción que me parece importante establecer es la que se da entre (i) tareas sobre la marcha (conocidas como tareas *online*) y (ii) tareas a posteriori (u *offline*). Esta distinción depende de si la medición de la respuesta se lleva a cabo mientras tiene lugar el proceso que subyace a la tarea que el participante realiza o se produce una vez terminado el proceso. La tarea de decisión léxica es una tarea a posteriori, mientras que otras tareas indirectas que examinaremos en el siguiente apartado, como la de detección de estímulos en el paradigma de atención dividida o las tareas de interferencia, son tareas que permiten obtener información sobre la marcha. Esta distinción cobra una gran importancia en la investigación sobre el procesamiento del lenguaje porque permite comprobar los cambios que se dan en el curso temporal de la tarea (y las operaciones subyacentes a la misma) y discriminar las reacciones espontáneas y automáticas que suscitan los estímulos en las primeras fases del procesamiento de las decisiones y estrategias conscientes del participante, que pueden encubrir o modificar esas reacciones espontáneas iniciales.

La tercera distinción que cabe destacar es la que corresponde a (i) tareas que se ejecutan bajo control voluntario del participante y (ii) tareas implícitas, que el sujeto realiza sin control consciente y voluntario. Las tareas que requieren juicios explícitos pertenecen a la primera categoría, pero se pueden combinar con tareas implícitas manipulando las condiciones de exposición de los estímulos y combinando aquellos que requieren la atención manifiesta del participante (para emitir un juicio sobre ellos) con otros que aparecen como elementos distractores o incluso se presentan de manera subliminal en el transcurso de la tarea. Aunque esta distinción es diferente a la anterior desde el punto de vista operacional, existe una estrecha relación entre ellas, ya que las tareas sobre la marcha inciden más en los procesos automáticos que no están sometidos al control voluntario del sujeto, mientras que las tareas a posteriori son, en general, más sensibles a las respuestas voluntarias y conscientes de los participantes.

Las siguientes distinciones dependen de factores externos a la propia tarea, desde los procesos específicos que son objeto de estudio hasta el tipo de población a la que pertenecen los participantes en la investigación. Así, podemos distinguir, en cuarto lugar, entre (i) tareas de percepción y/o comprensión del lenguaje y (ii) tareas de producción. Así mismo, como quinta distinción, podemos hablar de diferentes tareas de procesamiento del lenguaje según las diferentes modalidades de uso del mismo (auditiva-oral, escrita-visual o signada-visual). Sobre estas dos últimas distinciones, es importante puntualizar que los intereses de los lingüistas y de los psicólogos o los científicos cognitivos pueden ser divergentes en lo que respecta a las coincidencias o discrepancias entre los procesos de comprensión y producción del lenguaje o entre los procesos que tienen lugar bajo las distintas modalidades de externalización. Por una parte, los lingüistas pueden estar más inclinados a buscar las invariantes del procesamiento entre comprensión y producción o entre distintas modalidades de uso, mientras que, por otra, los científicos cognitivos tienden a indagar en las diferencias, que no tienen por qué ceñirse exclusivamente a aspectos superficiales o periféricos del procesamiento.

En otro orden de cosas, y de acuerdo con una sexta distinción, las tareas se deben ajustar a las características del participante, lo que nos lleva a diferenciar entre tareas diseñadas para adultos y tareas diseñadas para niños, o entre tareas en modo monolingüe y en modo bilingüe, cuando se mezclan o combinan lenguas distintas en la misma tarea, y, por último, tareas destinadas a evaluar a poblaciones especiales, como pacientes adultos con alteraciones neuropsicológicas de diversa índole (afasias, demencias) o personas

(generalmente niños, adolescentes o jóvenes) con alteraciones cognitivas o lingüísticas del desarrollo. Finalmente, un criterio importante en el diseño de una tarea, sea conductual o de otra especie, es el nivel de representación o de procesamiento o el componente del lenguaje que es materia de investigación, desde la fonología segmental y suprasegmental, la morfología, el léxico, la sintaxis, la semántica léxica y oracional, hasta la pragmática lingüística y el discurso.

Conviene puntualizar que las distinciones descritas en los últimos párrafos no son mutuamente independientes, y, por consiguiente, las clases de tareas que emergen de ellas se pueden cruzar unas con otras. Aquí se han presentado por separado con el propósito de exponer los criterios que rigen el diseño y la construcción de tareas en la investigación psicolingüística. No obstante, en la exposición de las tareas conductuales que sigue a este apartado, me centraré en tres de las distinciones descritas, a saber, las que se establecen entre tareas de comprensión y tareas de producción, entre tareas sobre la marcha y a posteriori, y entre tareas explícitas o controladas y tareas implícitas o no controladas. La combinación de estas tres distinciones dicotómicas se puede trasladar a un esquema en forma de espacio tridimensional, como el que se muestra en la Figura 1, en el que se podría localizar cada tarea individual bajo el supuesto de que cada una de las distinciones que se han establecido constituye una dimensión continua, en lugar de definir categorías disjuntas de tareas. En el próximo apartado presentaré las principales tareas conductuales que se emplean en la investigación empírica (sobre todo experimental) sobre el lenguaje siguiendo esta clasificación tripartita.

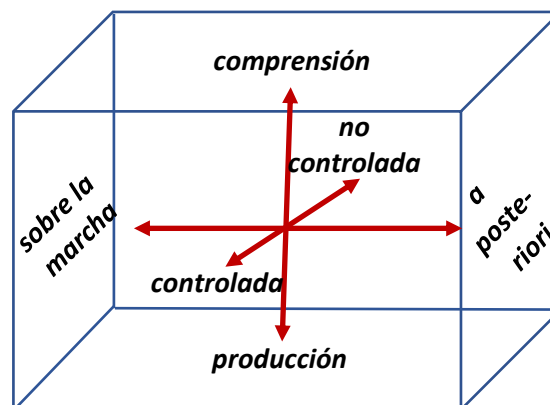


Figura 1. Tres dimensiones continuas bipolares para definir diferentes clases de tareas conductuales en la investigación sobre el procesamiento del lenguaje

3. Tareas conductuales en la investigación sobre el procesamiento del lenguaje

La noción de tarea, según se emplea en la investigación empírica, no se refiere solo a lo que los participantes hacen en una situación de recogida de datos, bien de forma espontánea o bajo las instrucciones del investigador, sino también a las condiciones bajo las que lo hacen, es decir, a los materiales o estímulos con los que se realiza y al modo en que estos se presentan en el transcurso de la misma. Otra advertencia que conviene hacer antes de proceder a la descripción de las tareas es que en los manuales y artículos sobre metodología de investigación en Psicolingüística, a menudo, se hace una distinción entre *tareas* y *paradigmas* de investigación. El término “paradigma” se utiliza en dos sentidos

muy diferentes en el marco de la metodología de investigación. En su acepción epistemológica, se refiere al conjunto o sistema de creencias que subyace a una metodología de investigación y que articula las hipótesis que emanan de ella (Guba y Lincoln 1994). En este sentido, en Psicología se suele hablar, por ejemplo, del “paradigma positivista” o del “paradigma constructivista” para hacer referencia a las raíces filosóficas de estas dos tradiciones de pensamiento y su impacto en los métodos de investigación. Sin embargo, en una acepción más pragmática, el término “paradigma” también hace referencia a una práctica o procedimiento que se emplea en la investigación experimental y que tiene por objeto establecer el modo en que se presentan los estímulos o materiales empleados en la tarea que realizan los participantes en un experimento. En este sentido, el concepto de “paradigma experimental” es transversal a la noción de tarea, lo que significa que un mismo paradigma puede utilizarse en combinación con distintas tareas.

Uno de los paradigmas experimentales más utilizados en Psicología es el paradigma de *priming*⁵, consistente en la presentación (bajo distintas formas y condiciones) de un estímulo, denominado “estímulo inductor” (*prime*, en inglés), que antecede al estímulo al que el participante debe responder, o “estímulo diana” (*target*, en inglés) (Bermeitinger 2014; Shao y Meyer 2017; Grey y Tagarelli 2018). El paradigma de *priming* permite manipular diferentes variables independientes de potencial interés en una investigación. La principal de ellas es la clase de relación entre el estímulo inductor y el estímulo diana. Cuando se utilizan palabras como estímulos, esta relación puede ser semántica (*delfín-ballena*), asociativa (*araña-tela*), morfológica (*niño-niña*), ortográfica y/o fonológica (*nueve-nuevo*). El paradigma de *priming* también se puede emplear con otros tipos de estímulos lingüísticos (oraciones) y con la combinación de imagen y palabra como estímulos inductor y diana en ambas direcciones. Una segunda variable importante que se suele manipular en este paradigma experimental es el desfase temporal entre los estímulos inductor y diana, una variable que puede medirse de dos maneras: mediante el “intervalo entre estímulos” (ISI, según las siglas de *inter-stimuli interval* en inglés), tiempo que media entre el final del estímulo inductor y el comienzo del estímulo diana; o bien mediante la “asincronía entre estímulos” (SOA, del inglés *stimulus onset asynchrony*), intervalo entre el inicio del estímulo inductor y el inicio del estímulo diana. Como veremos en los subapartados que siguen, el paradigma de *priming* se emplea tanto en tareas de comprensión como en tareas de producción del lenguaje.

Otra manipulación experimental que cabe definir como “paradigma” consiste en emplear dos tareas distintas que se deben ejecutar de forma simultánea, lo que obliga al sujeto experimental a administrar de manera eficiente sus recursos de atención y memoria a fin de poder realizar ambas con suficiencia. Por ello, este paradigma ha recibido el nombre de paradigma de *atención dividida* o de tarea dual (Nicol et al. 2006). Por lo común, una de las tareas (por ejemplo, reconocer una palabra o comprender un enunciado lingüístico previamente señalado) se considera principal, mientras que la otra es una tarea subsidiaria. Esta a su vez puede ser voluntaria (detectar o categorizar un estímulo distractor que aparece de forma inesperada) o involuntaria (escuchar pasivamente el estímulo distractor). El objetivo es comprobar el grado de interferencia que provoca el estímulo distractor al inducir la tarea subsidiaria. Asimismo, las dos tareas pueden realizarse en la misma modalidad sensorial –auditiva o visual– o en modalidades distintas. El paradigma de atención dividida permite incorporar al experimento las mismas variables independientes mencionadas en relación con el paradigma de *priming*.

3.1. Tareas conductuales en el estudio de la comprensión del lenguaje

La mayor parte de las tareas conductuales que se emplean en la investigación sobre los procesos de comprensión del lenguaje son tareas de juicios en las que se pide a los participantes que emitan una respuesta, generalmente dicotómica, acerca de una propiedad de los estímulos que se presentan, ya sean palabras aisladas, fragmentos de oraciones o enunciados completos. Esto permite registrar el tiempo de reacción ante el estímulo y, cuando existe un criterio de corrección de la respuesta, calcular la tasa de acierto en la tarea.

Entre las tareas de juicios más frecuentemente empleadas se cuentan la *tarea de decisión léxica*, los juicios semánticos sobre palabras (como la pertenencia de la palabra a un determinado dominio conceptual o categoría semántica) y distintas clases de juicios sobre oraciones, como la gramaticalidad o la plausibilidad. La tarea de decisión léxica es sensible a un amplio abanico de factores. Algunos de ellos son subléxicos, como la frecuencia de las sílabas y los bigramas que componen la palabra. Sin embargo, la mayoría son léxicos, entre los que cabe destacar la frecuencia de uso, la longitud de la palabra en letras o sonidos, su semejanza fonológica u ortográfica con otras palabras (los denominados “vecinos léxicos”), la edad estimada de adquisición, su composición morfológica, su categoría gramatical o aspectos de su significado. Por último, la tarea de decisión léxica también es sensible a factores contextuales, como la predictibilidad de la palabra en un contexto oracional o la “diversidad contextual”, es decir, el número de contextos distintos en los que puede aparecer la palabra (Igoa 2009).

El análisis de los efectos de los factores enumerados en el párrafo anterior muestra que, a mayor frecuencia de uso de sílabas, bigramas o palabras y a mayor predictibilidad contextual de una palabra, menor será el tiempo de respuesta en la tarea de decisión léxica. Por el contrario, las palabras largas y aquellas con mayor densidad de vecinos léxicos producen tiempos mayores de reconocimiento (Ballot et al. 2021). Por otra parte, la sensibilidad de esta tarea a factores contextuales y a las propiedades semánticas de las palabras indica que no solo garantiza el acceso al lexicón mental, sino que también implica la activación del significado de la palabra objeto de reconocimiento. En este aspecto, cabe destacar como factores facilitadores de la respuesta, por ejemplo, la “imaginabilidad” de la palabra (es decir, el grado en que se halla disponible una representación en imágenes del concepto denotado por la palabra), la diversidad contextual y la conectividad semántica de la palabra (esto es, el número de palabras con las que la palabra diana tiene relaciones directas de significado) (Adelman et al. 2006). Por último, cabe mencionar que todos estos factores pueden producir efectos múltiples, ya sea de forma aditiva, es decir, sumando el efecto independiente que ejerce cada factor, o interactiva, esto es, produciendo un efecto combinado de dos o más variables. Cabe añadir que los efectos pueden facilitar o entorpecer el proceso que se analiza y, con ello, el tiempo de reconocimiento de la palabra.

El único coste que conlleva el diseño de esta tarea es la necesidad de construir pseudopalabras, o secuencias de fonemas o letras que cumplan las reglas fonotácticas y ortotácticas de la lengua, pero que carezcan de significado, a fin de equilibrar la proporción de ensayos positivos y negativos⁶.

La tarea de decisión léxica contrasta, por el lado del significado, con tareas de juicios semánticos, como el de categorización, y por el lado de la forma, con la tarea de

denominación, repetición oral o lectura en voz alta de palabras. Aunque esta última tarea también asegura un contacto con el lexicón, parece ser menos sensible a factores contextuales y semánticos que la tarea de decisión léxica, pese a que tampoco está exenta de ellos (Wilson et al. 2013). Por tanto, la interpretación más común de los efectos diferenciales de ambas tareas es que la de decisión léxica pone en juego procesos post-léxicos como la integración de la palabra en el contexto oracional, un nivel que no alcanza la tarea de denominación de palabras (Igoa 2009).

Por su parte, las tareas de juicios son tareas a posteriori que permiten evaluar procesos no controlados y que permanecen por debajo del nivel de la conciencia en el reconocimiento y la comprensión de palabras. Una estrategia muy utilizada para este propósito es el uso de estas tareas bajo el paradigma de *priming*. El contraste relevante, en este caso, se da entre los efectos causados por el estímulo inductor cuando este es visible y cuando se expone en condiciones de enmascaramiento. Manipulando cuidadosamente el tiempo de exposición del estímulo inductor (en torno a 50 o 60 milésimas de segundo) y anteponiendo a su presentación un estímulo-máscara (por ejemplo, una fila de almohadillas o asteriscos) se ha conseguido generar efectos de facilitación en la tarea de decisión léxica con esta variedad de *priming* enmascarado (véase el esquema de la Figura 2). Esto ocurre así en el llamado *priming de repetición (casa-casa)* cuando hay semejanza ortográfica entre los estímulos inductor y diana (*caso-casa*) y también cuando ambos guardan una relación morfológica (*entiendo-entender*) o semántica (*marido-mujer*) (Forster y Davis 1984; Perea y Gotor 1997; Forster 1998; Sánchez-Casas et al. 2003). En cambio, cuando el estímulo inductor es una pseudopalabra (*cusa-casa*), el efecto de facilitación se desvanece (Lázaro et al. 2018) o incluso se hace inhibitorio, es decir, provoca un enlentecimiento de la respuesta y un mayor porcentaje de errores (Mathey et al. 2013).

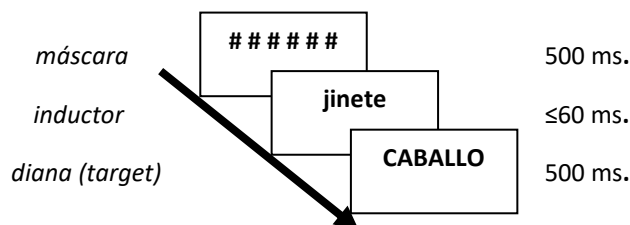


Figura 2. Secuencia de acontecimientos en un ensayo de *priming* semántico enmascarado, con los tres estímulos dispuestos en orden y sus respectivos tiempos de exposición

La obtención de efectos, ya sea facilitadores o inhibitorios, con el paradigma de *priming* requiere la comparación entre una condición en la que los estímulos inductor y diana están relacionados con otra, denominada “condición de control” o de “línea base”, en la que ambos estímulos no guardan ninguna relación. En el caso del *priming* enmascarado, la existencia de efectos de facilitación-sugiere, además, que el estímulo enmascarado se activa en la memoria del participante sin que este tenga conciencia de su presencia, produciendo una suerte de efecto de percepción subliminal. Esta activación encubierta se propaga al estímulo diana, reduciendo el tiempo de respuesta y aumentando la tasa de acierto en la tarea.

De todos los efectos referidos, el que ha generado un debate mayor entre los investigadores es el efecto de *priming* semántico enmascarado, debido a la relativa heterogeneidad de los materiales empleados a través de estudios diferentes en lo relativo a las relaciones semánticas entre los estímulos inductor y diana. Los autores más escépticos con la posibilidad de hallar efectos de facilitación semántica en condiciones de enmascaramiento arguyen que la noción de “relación semántica” se emplea con excesiva laxitud, pues a menudo comprende tanto relaciones semánticas estrictas (de hiperonimia/hiponimia, como en *insecto-hormiga*, entre palabras de la misma categoría conceptual, como *araña-hormiga*, o entre sinónimos o antónimos) como relaciones asociativas que no son rigurosa o exclusivamente semánticas; por ejemplo, pares como *perro-gato*, que mantiene una relación a la vez semántica y asociativa e incluso pares como *araña-tela* o *coche-carrera*, los cuales ni siquiera mantienen una relación semántica estricta y están únicamente asociados entre sí por el hecho de que han ocurrido juntos repetidas veces en la experiencia mundana del participante (Ferrand y New 2004). Sea como fuere, los efectos de *priming* semántico enmascarado con palabras obtenidos en la investigación no siempre son robustos y parecen ser sensibles a la naturaleza específica de la relación entre los miembros del par (Hutchison 2003).

Una tarea que se ha empleado abundantemente en la investigación sobre el análisis sintáctico (*parsing*) y la comprensión de enunciados y textos en la lectura es la *tarea de lectura autoadministrada*. Esta tarea consiste en la presentación de enunciados lingüísticos segmentados en regiones que el propio lector se administra sucesivamente (pulsando un botón) a medida que lee cada fragmento del enunciado (Mitchell y Green 1978; Jegerski 2014). Una variante de esta tarea es la denominada *tarea de escucha autoadministrada*, en la que los materiales se presentan por vía auditiva y el oyente se administra sucesivos fragmentos de la oración pulsando un botón para avanzar de uno a otro. El tiempo que media entre cada dos pulsaciones se registra como tiempo de lectura del segmento expuesto en la pantalla, lo que permite comparar los tiempos de lectura de los mismos segmentos bajo distintas condiciones experimentales, a fin de obtener una medida indirecta de la complejidad de procesamiento de los segmentos del enunciado. Esta tarea va normalmente acompañada de una pregunta de control situada al final de cada ensayo, a fin de asegurar que el lector presta atención al significado de la oración y no se limita a hacer una lectura mecánica de la misma.

Un caso muy conocido de aplicación de la tarea de lectura autoadministrada es su empleo en el estudio de los procesos de resolución de ambigüedades sintácticas en la adjunción de constituyentes, como se muestra en (1), en el que la cláusula de relativo (CR), señalada en cursiva, se puede adjuntar como modificador de cualquiera de los dos antecedentes nominales, identificados con subíndices (las barras indican la división de la oración en segmentos para la lectura y el segmento en negrita es el sintagma que desambigua la CR) (Cuetos y Mitchell 1988; Carreiras y Clifton 1993; Gilboy y Sopena 1996).

- (1) Alguien disparó / al criado₁ de la actriz₂ / que estaba en el balcón / con su marido.

Como se puede observar en el ejemplo, el segmento de la CR es totalmente ambiguo, ya que antes de presentar el sintagma final de la oración, se puede interpretar indistintamente que el criado o la actriz estaban en el balcón. La desambiguación se produce en el último segmento de forma pragmática (y, por tanto, probabilística), dado que, con esta

continuación, es más probable que fuera la actriz quien estaba en el balcón. En tales circunstancias, se puede estimar la preferencia de adjunción de la CR a partir del tiempo de lectura del segmento desambiguador, de tal modo que si la CR se adjunta al primer nombre –lo que se conoce como “adjunción alta”–, se produce un conflicto al leer el segmento desambiguador, pues este es más compatible con la adjunción de la CR al segundo nombre –lo que supone una “adjunción baja” de la CR.

La tarea de lectura autoadministrada se puede hacer efectiva en diversos formatos o condiciones de presentación de los segmentos: puede ser “acumulativa”, cuando los sucesivos segmentos van sumándose progresivamente al texto hasta completar el enunciado en su totalidad, o “no acumulativa”, en cuyo caso la aparición de cada nuevo segmento conlleva la desaparición del anterior (o su conversión en una fila de guiones), a fin de evitar movimientos oculares regresivos hacia porciones anteriores del texto. Esta forma de presentación no acumulativa se denomina procedimiento de “ventana móvil”, pues simula una ventana que se desplaza en el sentido de la lectura. Una forma alternativa de la presentación no acumulativa consiste en presentar cada segmento de la oración en el centro de la pantalla, de manera que cada nuevo segmento sustituya al anterior. Además de elegir la modalidad de presentación, el investigador que utiliza la tarea de lectura autoadministrada debe establecer el tamaño y características de las unidades en las que se han de segmentar las oraciones. Lo normal es utilizar criterios lingüísticos para dividir los estímulos, ya sea presentándolos palabra por palabra o sintagma por sintagma. En cualquier caso, lo más conveniente es que la comparación de los tiempos de lectura o procesamiento se efectúe entre condiciones lo más equivalentes posible, idealmente entre segmentos idénticos en contextos diferentes, como en las oraciones del ejemplo (2), en las que el adjetivo *caro* es el estímulo crítico precedido de una cópula distinta en cada caso (las barras representan la división entre segmentos).

- (2) a. El solomillo / de buey / era / caro ...
b. El solomillo / de buey / estaba / caro ...

Cuando esto no es posible, se debe procurar equiparar al máximo los estímulos que van a ser objeto de comparación, atendiendo a variables importantes como la longitud y la frecuencia de los estímulos. Una forma de remediar los inconvenientes derivados de la comparación de estímulos parcialmente distintos entre condiciones es calcular lo que se conoce como “tiempos residuales de lectura” de cada participante. Esta medida consiste en una estimación del promedio del tiempo de lectura de cada sujeto individual para cada estímulo del experimento en función de la longitud de los segmentos y del promedio del tiempo de lectura de cada sujeto individual en todos los ítems del experimento. Este tiempo estimado funciona como factor de corrección que se resta de los tiempos reales de lectura para suprimir o minimizar la variabilidad de los datos.

Además del problema que acabo de señalar, la lectura autoadministrada presenta otros inconvenientes. Al tratarse de una respuesta voluntaria, esta tarea está sujeta a sesgos y distorsiones provocados por decisiones estratégicas de los participantes, lo que constituye otra fuente de variabilidad en las respuestas. El problema de fondo es la escasa validez ecológica de la tarea, esto es, sus importantes diferencias con una tarea de lectura en condiciones naturales. Una precaución importante que hay que tomar a la hora de interpretar los tiempos de lectura en este tipo de tarea es que la carga de procesamiento asociada a un segmento dado se puede trasladar al siguiente, generando un efecto de

arrastre que tiende a engrosar el tiempo de respuesta de la región adyacente a la que se ha seleccionado como segmento crítico de la oración. El efecto de arrastre se agudiza especialmente al final de la oración, una región que conlleva una carga de procesamiento mayor porque en ella se tiende a recapitular y a evaluar la información analizada a lo largo de la oración. Por ello, no es aconsejable situar el segmento crítico en la posición final de la oración.

Uno de los principales inconvenientes de las tareas que he revisado hasta el momento en el ámbito de la comprensión del lenguaje es que se tratan de tareas a posteriori, dado que registran la respuesta del participante una vez que el proceso, que es objeto de análisis, ha concluido. Además, al tratarse de tareas que requieren una respuesta voluntaria, no se puede evitar la intrusión de estrategias del lector u oyente, como la búsqueda de un compromiso entre la rapidez y la precisión, mencionada en el apartado anterior. Estos problemas se pueden paliar recurriendo a tareas que requieren una respuesta no demorada, lo que permite evaluar sobre la marcha los procesos que son objeto de estudio. Las tareas que voy a describir seguidamente registran las respuestas de los participantes sobre la marcha y se emplean bajo el paradigma de “atención dividida”. Se trata de las tareas de *detección (o monitoreo) de estímulos* sobre la marcha y las tareas de *priming transmodal*. En ambas, el participante tiene como tarea principal la comprensión de una palabra u oración presentada por vía auditiva, a la vez que ha de prestar atención a un estímulo subsidiario, un distractor que aparece de forma imprevista y requiere una respuesta inmediata, de ahí su caracterización como tareas de atención dividida.

La tarea de detección de estímulos fue una tarea muy utilizada en Psicolingüística desde los años 60 del pasado siglo (Abrams y Bever 1969) hasta el advenimiento de nuevas tecnologías de recogida de datos, como el registro de potenciales evocados o de movimientos y fijaciones oculares en la lectura, que permitieron diseñar tareas sobre la marcha utilizando respuestas que no requieren un control consciente y voluntario por parte del sujeto. El estímulo subsidiario empleado en esta tarea puede ser un fonema, un ruido (chasquido, clic o tono de muy breve duración) que se inserta en un lugar de la oración, o incluso una palabra. El participante ha de responder pulsando un botón en cuanto detecte el estímulo distractor. La latencia de la respuesta de detección se interpreta como indicio de la carga de procesamiento de la oración en el dominio en el que se encuentra el distractor, de tal modo que una latencia más alta revela un mayor consumo de recursos cognitivos destinados al análisis lingüístico que se sustraen a la tarea de detección. Esta tarea permite manipular la localización del estímulo distractor y comparar la carga de procesamiento en momentos distintos del enunciado. Aunque es una tarea poco natural y basada en una respuesta manual voluntaria, la tarea de detección de estímulos permite obtener medidas sobre la marcha de los procesos de comprensión, lo que hace aconsejable su utilización en combinación con otras tareas a posteriori que se emplean en este ámbito de investigación. Uno de los usos más frecuentes de esta tarea ha sido el de evaluar las demandas variables de memoria en regiones específicas de la oración (Mota e Igoa 2017). Así, en el transcurso del análisis sintáctico de ciertas oraciones es necesario mantener activado el núcleo nominal del sintagma de sujeto para establecer la concordancia gramatical con un verbo que está por llegar, como ocurre en la posición señalada por la flecha en la oración (3a), en contraste con la misma posición en la oración (3b), donde la concordancia ya ha sido establecida y, por consiguiente, la carga de memoria es hipotéticamente menor.

- (3) a. El conductor vio que la rueda del remolque del camión ↓ de mudanzas estaba pinchada
- b. El conductor cambió la rueda del remolque del camión ↓ de mudanzas

Un procedimiento algo más complejo, pero bastante utilizado en la investigación psicolingüística, es el paradigma de *priming* transmodal. Este paradigma es similar a la tarea de detección porque también requiere simultaneizar dos tareas, la comprensión de una oración hablada y una respuesta a otro estímulo lingüístico. De esta manera, esta tarea consiste normalmente en una decisión léxica o un juicio semántico a una palabra o pseudopalabra que se presenta en modalidad visual en un punto crítico de la oración y sincronizada con el estímulo principal. La palabra o frase que se presenta por vía auditiva actúa como estímulo inductor que desencadena el efecto de *priming* sobre el estímulo “diana” visual que está sincronizado con él. El uso de dos modalidades sensoriales, la auditiva y la visual, hace de este paradigma un procedimiento transmodal. El *priming* transmodal se considera un procedimiento apropiado para evaluar procesos de activación de información latente (representaciones lingüísticas inconscientes) que interviene en el análisis y la interpretación de un enunciado (Roberts 2014). Por ejemplo, en situaciones de ambigüedad léxica, se plantea la cuestión de cuándo se accede a la información del contexto para determinar el significado de la pieza léxica ambigua y, en relación con ello, hasta qué punto (y en qué momento del proceso) se activan los significados alternativos que no son congruentes con el contexto en el que aparece la palabra en cuestión. Una manera de indagar en este proceso es sincronizando la presentación visual de una palabra (p.ej., *flor* o *piso*) asociada con uno de los sentidos de una palabra ambigua (p.ej. *planta*) que aparece en un contexto de frase o discurso presentado auditivamente, que a su vez favorece uno de los significados del término ambiguo (p.ej. un contexto que habla de *jardines, macetas, abono, riego*). En estas circunstancias, se pide a los participantes que emitan un juicio acerca de la palabra presentada visualmente. Esto permite comprobar si el tiempo de reconocimiento de la palabra escrita (*flor* o *piso*) difiere en función de la congruencia con el significado de la palabra ambigua favorecido por el contexto (una planta como vegetal) (Swinney 1979). Lo mismo que en la tarea de detección, esta permite seleccionar el momento temporal exacto en el que se sincronizan las entradas auditiva y visual, y con ello explorar las variaciones temporales que pueda haber en el nivel de activación de los distintos significados de la palabra ambigua. Los resultados de estos estudios indican que en el momento de oír la palabra ambigua se activan todos sus significados, con independencia de cuál sea congruente con el contexto previo. No obstante, este efecto es transitorio, ya que el significado apropiado al contexto (p. ej. *planta* vegetal) se impone poco después, inhibiendo los sentidos contextualmente incongruentes (p. ej. *planta* de edificio).

La técnica de *priming* transmodal se ha empleado también para estudiar el estado de activación de las huellas sintácticas generadas por el movimiento de constituyentes en frases interrogativas o cláusulas de relativo, según postulan algunas teorías lingüísticas (Nicol y Swinney 1989). En algunos de estos estudios se han empleado estímulos visuales (dibujos) asociados al antecedente nominal de una huella sintáctica que aparecería en una posición diferida de la estructura en una oración presentada en modalidad auditiva. Por ejemplo, en la oración del ejemplo (4) se postula una huella sintáctica (h_i) en la posición [2] que hay que enlazar en una cadena que la une con el pronombre relativo que encabeza la cláusula subordinada (que_i) y con el antecedente que lo precede (*ardilla*).

- (4) Alfredo persiguió a la ardilla a la que_i el simpático mono explicó las complicadas [1] reglas del juego h_i [2] en la clase del miércoles pasado

Para comprobar esto, se mide el tiempo de respuesta a la pregunta de si un dibujo que aparecerá en un momento dado durante la oración que van a oír es un ser vivo. Se trata de comparar el tiempo de respuesta a un dibujo que requiere una respuesta afirmativa a la pregunta (una ardilla es un ser vivo), en contraste con otro que requiere una respuesta negativa (un cepillo de dientes), en dos lugares distintos de la oración (posiciones [1] y [2]). La hipótesis es que si hay una huella sintáctica en la posición [2], la respuesta al dibujo de la ardilla será más rápida en dicha posición que al dibujo del cepillo de dientes, mientras que en la posición [1] no habrá diferencias en la respuesta a ambos dibujos, ya que ahí no tiene lugar la reactivación del antecedente (Felser y Roberts 2007).

En el presente apartado, he pasado revista a algunas de las tareas conductuales más utilizadas en la investigación sobre los procesos de reconocimiento de palabras y comprensión de palabras y oraciones, destacando el empleo de los paradigmas de *priming* y atención dividida en combinación con algunas de estas tareas. El hecho de que algunas de estas tareas, como la de denominación, incluyan un componente de producción no impide que sean clasificadas como tareas de comprensión, ya que en estas tareas la producción de la respuesta se emplea como indicio para contrastar hipótesis y obtener información sobre los procesos de reconocimiento o comprensión. Por otra parte, es preciso tener en cuenta que estas tareas se acoplan regularmente con métodos no conductuales de recogida de datos, como los registros de movimientos oculares y de potenciales evocados, o técnicas de obtención de imágenes cerebrales como la resonancia magnética funcional o la magnetoencefalografía. Por ello, la división entre tareas conductuales y otras técnicas no conductuales en el ámbito de la Psicolingüística experimental tiene utilidad a título descriptivo, pero no representa una división sustantiva entre métodos alternativos de investigación empírica. En el próximo apartado, me ocuparé de describir algunas tareas conductuales utilizadas en la investigación sobre la producción del lenguaje, un territorio menos explorado en términos de investigación empírica que el de la comprensión del lenguaje, tanto en el pasado como en el presente.

3.2. Tareas conductuales en el estudio de la producción del lenguaje

Como es natural, las tareas conductuales que se emplean en la investigación de la producción del lenguaje van orientadas a hacer hablar a los participantes para observar el producto de esta tarea y diversos parámetros asociados a ella, como el tiempo que tardan en iniciar la respuesta verbal o los errores que cometen en la producción. En este sentido, las tareas de producción se sitúan en un continuo de espontaneidad o naturalidad que es inversamente proporcional al grado de manipulación de las condiciones bajo las que se produce el comportamiento verbal, desde tareas totalmente espontáneas, donde la manipulación es inexistente y los grados de libertad indefinidos, hasta tareas muy controladas, en las que se manipulan distintas variables que inciden en el comportamiento verbal y los grados de libertad están mucho más restringidos. En el extremo más natural de este continuo se encuentra el habla espontánea, de la que se registran algunos fenómenos que la acompañan, principalmente los errores (o *lapsus linguae*) que cometen los hablantes (Garrett 1975; Del Viso 2002; Harley 2006) y las disfluencias (falsos comienzos, titubeos, pausas o el conocido fenómeno de la “punta de la lengua”) que se producen en el habla espontánea (Butterworth 1980; González 1996). El papel del

investigador que se enfrenta a estos datos se reduce a seleccionar las muestras de habla y aplicar procedimientos de análisis de dichas muestras de acuerdo con principios y criterios derivados de las teorías y modelos al uso. En el extremo más controlado hallamos tareas de repetición de palabras y oraciones habladas o de lectura en voz alta, en las que el/la hablante suele tener poco margen de maniobra y el investigador puede controlar al máximo los materiales que se emplean en la tarea. En lo que sigue, voy a describir brevemente dos ejemplos de tareas, cada una de las cuales ocupa un lugar opuesto en los extremos del continuo, para después ocuparme de otras tareas conductuales que se encuentran en lugares intermedios.

El estudio de los errores espontáneos del habla ha atraído la atención de psicólogos y lingüistas desde hace mucho tiempo. Estas piezas de comportamiento, que obedecen a un fallo transitorio y parcial de los mecanismos de planificación de mensajes hablados y de su formulación lingüística, y no a un menoscabo en el conocimiento de la gramática o del vocabulario, se han examinado con diversos propósitos. El que nos interesa aquí es el análisis de las restricciones y regularidades que exhiben, habida cuenta de que cuando nos equivocamos al hablar, lo hacemos de manera sistemática y predecible, de acuerdo con unas pautas definidas. El examen de estas regularidades permite reconstruir las operaciones que subyacen a la producción del habla. Así, por ejemplo, hay dos grandes categorías de error que afectan, respectivamente, a la selección de unidades lingüísticas (sintagmas, palabras, morfemas, sílabas o fonemas), lo que se traduce en errores de sustitución, mezcla, adición u omisión de estas unidades, o bien a la ordenación de las mismas en secuencias o combinaciones, lo que puede provocar desplazamientos, trasposiciones o repeticiones de elementos que han sido correctamente seleccionados en el plan del enunciado. Una regularidad que exhiben los errores de esta última clase es que las unidades que se ven afectadas en ellos pertenecen a la misma categoría. Así, por ejemplo, cuando se produce una trasposición o intercambio de dos palabras o morfemas en la oración, ambos elementos pertenecen a la misma clase gramatical (palabras de contenido o funcionales, y dentro de cada clase, a nombres, verbos, adjetivos, pronombres o preposiciones, morfemas flexivos o derivativos, prefijos o sufijos).

Estas pruebas, como otras que se observan en aspectos más específicos del proceso de producción verbal, revelan que los mecanismos de producción del lenguaje se ajustan a unas reglas implícitas que se aplican en un orden sistemático en tiempo real durante la planificación y ejecución del enunciado. Sin embargo, como sucede con cualquier método de observación del comportamiento espontáneo, el estudio de los errores presenta algunos inconvenientes, que se derivan, por un lado, de los sesgos perceptivos del observador y del hecho de que unos errores sean más salientes o perceptibles que otros, y, por otro, de la dificultad que surge a menudo a la hora de clasificar errores en categorías discretas establecidas a priori, dada la ambigüedad inherente que presentan muchos de ellos. Estos problemas han llevado a algunos investigadores a someter el estudio de los errores y de otras disfluencias del habla a un control experimental mediante la estrategia de provocarlos artificialmente en el laboratorio (Baars y MacKay 1978). Aunque esta metodología es, en general, poco productiva, pues arroja una proporción de errores apenas superior al 5 o el 10 % de los ensayos, ha permitido explorar de manera sistemática y precisa algunos mecanismos de la producción, como es el caso del establecimiento de concordancia gramatical (sujeto-verbo en número, o entre determinante, nombre y adjetivo, dentro de un SN, en género y/o número), un proceso que ha demostrado ser sensible a errores de “atracción” como los que se explican a continuación. En

experimentos en los que se emplea una tarea de completar oraciones (con una cópula o con un verbo previamente indicado), se ha puesto de manifiesto que la interposición de un sintagma nominal en plural como complemento del nombre del núcleo singular de sujeto y el verbo provoca un fallo en la asignación del rasgo de número al verbo que produce el hablante para completar la oración, lo que origina un error de atracción de número del nombre más cercano (en plural) al verbo, como en el ejemplo (5) (Vigliocco et al. 1995), en el que el plural de *botellas* es atraído por la cópula que le sigue.

(5) La etiqueta de las botellas ... estaban manchadas

En el extremo opuesto del continuo espontáneo-controlado en las tareas de producción, se encuentran las tareas de repetición o lectura en voz alta de enunciados. A pesar de que, en apariencia, la repetición de una frase es una tarea sencilla y superficial que apenas consume recursos cognitivos, en realidad es una tarea reconstructiva, dado que requiere una reformulación lingüística en toda regla que abarca desde el significado del enunciado hasta la preparación y ejecución de un plan fonético. Naturalmente, esta tarea varía en su grado de dificultad en función de las demandas de memoria que impone, que dependen sobre todo de la longitud y la complejidad estructural del material. Por esta razón, la repetición de frases se ha empleado con frecuencia en poblaciones “especiales”, como niños en fase de adquisición de la lengua nativa, aprendices de una segunda lengua o personas con alteraciones del lenguaje. Así pues, la tarea de repetición es más una tarea que afecta a la memoria operativa que a la memoria episódica (o recuerdo textual o literal de frases). Una variante de esta tarea, que otorga más grados de libertad al participante, consiste en suministrarle un material lingüístico hablado y solicitarle que lo reconstruya a partir de una pregunta que se formula sobre dicho material y bajo unas instrucciones concretas (Fernández 2005). De ello se ofrece un ejemplo en (6).

- (6) Investigador: *El invitado impresionó al hermano del novio. ¿Qué novio?
El novio que (a menudo inconscientemente) roncaba.*
Participante: *El invitado impresionó al hermano del novio que (a menudo
inconscientemente) roncaba.*

Este tipo de tarea se ha empleado con gran provecho en los estudios sobre producción de la prosodia oracional, especialmente en relación con la función demarcativa de la prosodia, que facilita la agrupación de los constituyentes oracionales y determina las dependencias gramaticales entre ellos. En este sentido, la prosodia se revela como un factor de desambiguación oracional y la elección de un perfil de entonación puede indicar las preferencias de interpretación del enunciado por parte del hablante, sus “intenciones comunicativas” tal y como estas se plasman en la producción del enunciado. Así, en el ejemplo (6), la oración producida por el participante, que es ambigua en su versión escrita, quedará desambiguada a través de la prosodia, mediante la manipulación de los tonemas⁷ y las pausas en lugares críticos del enunciado ([al hermano] [del novio]).

El empleo del método experimental en los estudios de la producción del lenguaje ha encontrado en el paradigma de *priming* un recurso idóneo para manipular las condiciones bajo las cuales se generan enunciados y, en consecuencia, para restringir el espacio de decisiones en que los hablantes se mueven a la hora de producir un enunciado verbal. De hecho, alguna de las tareas presentadas en este apartado se pueden considerar casos del

uso de *priming* en producción, como por ejemplo la tarea de completar oraciones con un atractor de número (véase el ejemplo (5)), en la que el morfema de plural de la palabra *botellas* sirve de estímulo inductor que puede condicionar la respuesta del hablante. En otras tareas de producción se utiliza la técnica de *priming* de manera distinta. Una de las tareas más fértiles en este sentido es la que se conoce como *paradigma de interferencia dibujo-palabra* (Glaser 1992; Levelt et al. 1999; Farrell et al. 2012; Jescheniak et al. 2020). En ella se combina una tarea de denominación de objetos (mediante dibujos) con un procedimiento de *priming* léxico: al participante se le presenta un dibujo que tiene que nombrar (por ejemplo, una mesa) y, de forma sincronizada con el dibujo, se expone una palabra escrita sobreimpuesta en el dibujo (o pronunciada mientras el dibujo está en pantalla). La palabra inductora, o estímulo distractor, puede guardar con el nombre del dibujo una relación de significado (*silla-mesa*), una relación de forma o pronunciación (*meta-mesa*) o ninguna relación en particular (*cara-mesa*). La tarea del sujeto es simplemente nombrar el dibujo. Con esta presentación se pretende evaluar el efecto que produce la palabra distractora en el tiempo de respuesta y el acierto en dar con el nombre correcto. A esta manipulación de los estímulos, se le añade la de la sincronía entre el dibujo y la palabra, de modo que el dibujo puede anteceder a la palabra, la palabra al dibujo o ambos pueden ser presentados sincrónicamente, oscilando el intervalo entre ambos sucesos desde -300 a +300 milisegundos, donde el extremo negativo corresponde a la secuencia palabra-dibujo y el positivo, a la secuencia dibujo-palabra.

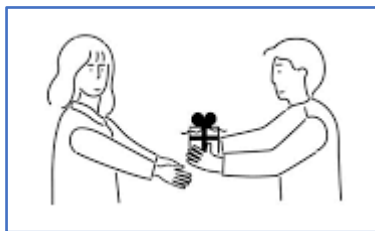
El objetivo de estos experimentos es poner a prueba la hipótesis de que el acceso al significado y el acceso a la forma del nombre del dibujo constituyen etapas segregadas en el proceso de recuperación de palabras del lexicón en la tarea de denominación y que, por tanto, la arquitectura del componente léxico en la producción presenta una estructura bipartita. El efecto más comúnmente hallado en los experimentos de interferencia-dibujo-palabra es un efecto inhibitorio sobre el tiempo de respuesta de nombrar el dibujo cuando la palabra se presenta antes que el dibujo, pero solo si la relación entre ella y el nombre del dibujo es semántica, y un efecto facilitador cuando el dibujo antecede a la palabra, pero solo cuando existe una semejanza de forma entre ambos nombres. Además, por regla general, estos efectos contrapuestos no se solapan en el tiempo. La explicación que se ofrece de estos resultados es que los conceptos lexicalizados (o expresables mediante piezas léxicas individuales) compiten entre sí, lo que, en nuestro ejemplo, origina una inhibición del concepto ‘mesa’ por la activación previa del concepto ‘silla’ al presentar la palabra antes que el dibujo. En cambio, las formas léxicas /*meta*/ y /*mesa*/ se activan mutuamente a causa de los sonidos que comparten, de modo que la activación de los fonemas (m/, /e/ y /a/ de /*meta*/), que también forman parte de la palabra *mesa*, ejercen un efecto excitatorio en una fase tardía del proceso, una vez activado el concepto ‘mesa’ (al que, recordemos, se accede primero a partir de la presentación del dibujo) y estando en marcha la búsqueda de la forma fonológica que le corresponde.

El paradigma de *priming* también se ha utilizado con bastante éxito en el estudio de los procesos de codificación gramatical en la producción de oraciones, aprovechando un fenómeno muy asentado y estable que se ha observado fuera del laboratorio, como es la tendencia a repetir la estructura sintáctica de oraciones previamente escuchadas o leídas en enunciados que el hablante emite posteriormente. Esta tendencia a la repetición de estructuras, que se observa con demoras muy largas, de incluso días o semanas, entre el estímulo escuchado o leído (el inductor) y el enunciado en el que se repite su estructura (la oración “diana”), se conoce como efecto de *priming sintáctico* (Bock 1986a, 1996;

Mahowald et al. 2016). Se suele interpretar apelando a la fortaleza de una huella que se crea en la memoria episódica y que facilita una mayor economía de recursos en la planificación de oraciones. En la reproducción de este fenómeno en el laboratorio, se presenta una oración inductora con una determinada estructura sintáctica y orden de constituyentes, como las del ejemplo (7), y seguidamente se expone una imagen de un suceso que los hablantes tienen que describir en una sola oración, como el dibujo en (8). Como se puede apreciar en los ejemplos, la estructura argumental de las oraciones en (7) es la misma que la del evento representado en (8), una estructura con tres argumentos que se expresa mediante una oración ditransitiva. Sin embargo, el orden de constituyentes (en concreto, de los objetos directo e indirecto del verbo) varía en la descripción verbal que se puede hacer de la escena. El efecto de *priming* se reflejaría en la tendencia a copiar la estructura de la oración inductora en la descripción verbal del dibujo, como sucede en los ejemplos en (9), cuyas variantes (a) y (b) se emparejan con las mismas variantes en (7).

- (7) a. Un guardia le puso a un automovilista una multa
b. Un guardia le puso una multa a un automovilista

(8)



- (9) a. El chico le está dando/entregando a la chica un regalo
b. El chico le está dando/entregando un regalo a la chica

Hay dos aspectos que se deben destacar del efecto de *priming* sintáctico: en primer lugar, que el efecto es genuinamente sintáctico, esto es, emerge de la configuración de los constituyentes (en el caso del ejemplo, el orden en que aparecen), y no de otras características más superficiales de la oración, como el uso de artículos, preposiciones o nombres concretos. Como se puede apreciar en los ejemplos anteriores, el efecto de *priming* es sensible al orden de los constituyentes. El carácter no superficial del nivel en el que se produce el efecto se confirma al observar que oraciones inductoras como las de (10), que presentan diferencias superficiales (la preposición *hacia*) producen el mismo efecto que las de (7) en la respuesta verbal al dibujo en (8).

- (10)a. La viuda condujo hacia la iglesia su viejo Mercedes
b. La viuda condujo su viejo Mercedes hacia la iglesia

El ejemplo (10) también deja patente que el efecto de *priming* sintáctico no se puede atribuir a la repetición de la estructura temática del enunciado, pues las oraciones en los ejemplos (9) y (10) no comparten la misma estructura argumental. Otro dato de interés sobre este efecto es que tampoco depende de la repetición de palabras específicas, si bien es cierto que queda reforzado cuando se repiten algunas piezas léxicas entre las oraciones inductoras y los nombres de un objeto o participante o el propio verbo representado en el dibujo (por ejemplo, con una oración inductora como *el policía dio un puñetazo al manifestante*). En conclusión, las propiedades del *priming* estructural favorecen un nivel

de representación sintáctica único, independiente de la semántica y relativamente abstracto, pero con constituyentes ordenados.

Los efectos de *priming* en la producción del lenguaje abundan por doquier. Se han obtenido efectos de facilitación por *priming* oracional (semántico) en la producción de palabras, utilizando frases como estímulos inductores y palabras como estímulos diana. Así, en una tarea de denominación de dibujos, el contexto oracional facilita la producción de palabras tanto semántica como fonológicamente relacionadas con el estímulo diana (el nombre del dibujo), generando un efecto de competición. También se han observado efectos de *priming* léxico en la producción de oraciones: cuando se presenta una palabra semánticamente relacionada con el nombre de uno de los personajes que aparecen en un dibujo, los hablantes tienden a mencionar, en primer lugar, a ese personaje, alterando la estructura sintáctica de la oración. Así, por ejemplo, produciendo una oración en voz pasiva, en lugar de activa, en oraciones transitivas (Bock 1986b). Por último, recordemos que el *priming* fonológico produce efectos de facilitación en tareas de denominación de dibujos, según se ha visto en los experimentos que usan el paradigma de interferencia dibujo-palabra mencionados. Pero este efecto facilitador de la fonología también parece afectar favorablemente a los estados de punta de la lengua, poniendo de relieve que los asociados fonológicos de una palabra (sean palabras o pseudopalabras) pueden ayudar a reducir o suprimir el bloqueo que se produce en estas interrupciones transitorias del habla.

4. Conclusiones

En el presente artículo se han expuesto los principales métodos de recogida de datos empíricos empleados en la investigación sobre el procesamiento del lenguaje que caen bajo la rúbrica de “tareas conductuales”. De entre los diversos criterios que se pueden utilizar para clasificar estos métodos, se ha establecido una distinción entre las tareas comúnmente empleadas en el estudio de los procesos de comprensión y las utilizadas en la investigación sobre la producción del lenguaje. Asimismo, se ha enfatizado la noción de paradigma, o procedimiento que sirve para guiar la aplicación de las tareas expuestas al estudio de problemas concretos, algunos de los cuales han servido de ejemplo en la exposición de las tareas examinadas.

Las tareas conductuales son necesarias en cualquier clase de investigación empírica, aunque solo sea porque si queremos recoger datos sobre el comportamiento de las personas, es imprescindible que actúen, que respondan de algún modo a alguna clase de estímulos. El escepticismo y las críticas que suscitan las tareas conductuales son en parte comprensibles, aunque quizá un tanto innecesarias. Son comprensibles porque adolecen de algunas limitaciones que no afectan a otras tareas más prestigiosas y utilizadas en la actualidad, como las respuestas electrofisiológicas, las imágenes cerebrales o los movimientos oculares, por citar tres ejemplos bien conocidos. En este artículo he señalado varias veces que uno de los problemas de las tareas conductuales es su dependencia de respuestas voluntarias y de decisiones conscientes de los participantes. Esto las hace susceptibles a la intrusión de factores estratégicos, de criterios, de cálculos y de expectativas que pueden utilizar los sujetos al ejecutar su respuesta y que contribuyen a ocultar o distorsionar procesos hipotéticamente más puros o prístinos que acontecen de manera automática y por debajo del nivel de la conciencia y que, al decir de algunos, deberían ser el verdadero objeto de una investigación libre de prejuicios. Un segundo

problema de las tareas conductuales, en su dimensión cuantitativa, es su dependencia de las medidas de tiempo, medidas que no hablan por sí solas, sino que han de ser interpretadas y ofrecen información limitada, pues no alcanzan más allá de la idea de que el tiempo es un indicador de la carga o la complejidad del procesamiento.

Hasta aquí los principales inconvenientes. No obstante, incluso reconociendo estos defectos, considero que las tareas conductuales han recibido un trato inmerecido, ya que las críticas que han recibido se pueden dirigir también a otras tareas más sofisticadas. Ninguna tarea realizada por humanos adultos permite garantizar un acceso privilegiado a los procesos o fenómenos que pretende explicar. Es cierto que hay procedimientos de recogida de datos que no requieren respuestas voluntarias. Sin embargo, estas tareas tampoco ofrecen soluciones plenamente satisfactorias a las necesidades del investigador. Todas requieren inferencias y parten de supuestos aceptados convencionalmente por la comunidad científica y raramente discutidos. Esta no es una idea nueva ni original, sino una simple constatación del carácter siempre indirecto y constructivo de la investigación científica, incluida la lingüística. En todo caso, las tareas conductuales siempre están presentes y se utilizan regularmente, bien solas o en combinación con otros procedimientos de recogida de datos. Esta es una de las razones por las que la división de las tareas en conductuales y de otra naturaleza no es más que una estrategia conveniente que utilizamos a efectos expositivos, pero no plantea una alternativa que obligue al científico a elegir entre distintas opciones.

La elección de una tarea, conductual o de otra clase, depende de los objetivos de la investigación, de las hipótesis que guían la indagación y del diseño que construimos y las variables que elegimos para contrastar esas hipótesis. Todo eso, a su vez, depende del soporte teórico en el que se asientan los objetivos y las hipótesis de la investigación. Una creencia que ha adquirido gran relevancia en la metodología de investigación en la ciencia cognitiva y en la Psicolingüística experimental es la idea de que las tareas sobre la marcha ofrecen soluciones más adecuadas y eficaces a los problemas de la investigación y, por ello, tienen más crédito que las tareas a posteriori. Pero lo cierto es que eso depende del problema que es objeto de investigación. ¿Por qué, entonces, son más apreciadas las tareas sobre la marcha? ¿Acaso porque son las que mejor permiten examinar las representaciones lingüísticas no accesibles a la conciencia? En mi opinión, no es así. En este artículo he tratado de mostrar que las tareas conductuales disponen de recursos adecuados y suficientes para explorar procesos y representaciones implícitos. Los paradigmas de *priming* expuestos a lo largo de estas páginas son buena muestra de ello. Por eso, un criterio de clasificación de las tareas expuestas en este artículo que cabría considerar igualmente significativo es el que establece una división entre tareas diseñadas para evaluar procesos controlados y tareas que permiten rastrear procesos automáticos. En cualquier caso, no es cuestión de elegir, pues, a fin de cuentas, cada tarea de las presentadas aquí, como en otros artículos de este monográfico, tiene su utilidad.

5. Referencias

- Abrams, Kenneth; Bever, Thomas G. 1969. Syntactic structure modifies attention during speech perception and recognition. *The Quarterly Journal of Experimental Psychology* 21.3: 280-290. DOI: 10.1080/14640746908400223
- Adelman, James S.; Brown, Gordon. D.; Quesada, José F. 2006. Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science* 17: 814-823.
- Baars, Bernard J.; MacKay, Donald G. 1978. Experimentally eliciting phonetic and sentential speech errors: Methods, implications, and work in progress. *Language in Society* 7.1: 105-109. DOI: 10.1017/S0047404500005352
- Ballot, Claire; Mathey, Stéphanie; Robert, Cristelle. 2021. Word imageability and orthographic neighbourhood effects on memory: a study in free recall and recognition, *Memory* 29.6: 829-834. DOI: 10.1080/09658211.2021.1921216
- Balota, David A.; Chumbley, James I. 1984. Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance* 10: 340-357. DOI: 10.1037//0096-1523.10.3.340
- Bermeitinger, Christina. 2014. Priming. En Z. Jin, ed. *Exploring implicit cognition: Learning, memory, and social cognitive processes*. Hershey, PA: Information Science Publishing, pp. 16-60.
- Berwick, Robert; C., Friederici, Angela D.; Chomsky, Noam; Bolhuis, Johan J. 2012. Evolution, brain, and the nature of language. *Trends in Cognitive Sciences* 17.2: 89-98. DOI: 10.1016/j.tics.2012.12.002.
- Bock, J. Kathryn. 1986a. Syntactic persistence in language production. *Cognitive Psychology* 18: 355-387. DOI: 10.1016/0010-0285(86)90004-6
- Bock J. Kathryn. 1986b. Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 12.4: 575-586. DOI: 10.1037/0278-7393.12.4.575
- Bock, J. Kathryn. 1996. Language production: Methods and methodologies. *Psychonomic Bulletin and Review* 3.4: 395-421. DOI: 10.3758/BF03214545
- Butterworth, Brian. 1980. Evidence from pauses in speech. En B. Butterworth, ed. *Speech Production*, Vol. I. New York: Academic Press, pp. 154-176.
- Carreiras, Manuel; Clifton, Charles. 1993. Relative clause interpretation preferences in Spanish and English. *Language and Speech* 36.4: 353-372.
- Christiansen, Morten H.; Chater, Nick. 2016. *Creating language: Integrating evolution, acquisition, and processing*. Cambridge, MA: The MIT Press.
- Cuetos, Fernando; Mitchell, Don C. 1988. Crosslinguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition* 30: 73-105. DOI: 10.1016/0010-0277(88)90004-2
- Eguren, Luis; Fernández Soriano, Olga. 2004. *Introducción a una sintaxis minimalista*. Madrid: Gredos.
- Ernestus, Mirjam; Cutler, Anne. 2015. BALDEY: A database of auditory lexical decisions, *The Quarterly Journal of Experimental Psychology* 68.8: 1469-1488. DOI: 10.1080/17470218.2014.984730
- Farrell, Meagan T.; Abrams, Lise; & White, Katherine K. 2012. The role of priming in lexical access and speech production. En N. Hsu y Z. Schütt, eds. *Perspectives on*

- cognitive psychology. Psychology of priming*. Nova Science Publishers, pp. 205-244.
- Felser, Claudia; Roberts, Leah. 2007. Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research* 23.1: 9-36.
- Fernández, Eva. 2005. The prosody produced by Spanish-English bilinguals: A preliminary investigation and implications for sentence processing. *Revista da Abralín* 1.2: 109-141.
- Ferrand, Ludovic; New, Boris. 2004. Semantic and associative priming in the mental lexicon. In P. Bonin, ed. *Mental lexicon: "Some words to talk about words"*. Nova Science Publishers, pp. 25-43.
- Ferreira, Fernanda; Patson, Nikole D. 2007. The 'good enough' approach to language comprehension. *Language and Linguistics Compass* 1.1-2: 71-83. DOI: 10.1111/j.1749-818x.2007.00007.x
- Fodor, Janet D.; Inoue, Atsu. 1994. The diagnosis and cure of garden-paths. *Journal of Psycholinguistic Research* 23: 407-434. DOI: 10.1177/00238309000430030201
- Fodor, Janet D.; Nickels, Stefanie; Schott, Esther. 2017. Center-embedded sentences: What's pronounceable is comprehensible. En R.G. de Almeida y L.R. Gleitman, eds. *On concepts, modules, and language: Cognitive science at its core*. Oxford, R.U.: Oxford University Press. DOI: 10.1093/oso/9780190464783.001.0001
- Forster, Kenneth I. 1998. The pros and cons of masked priming. *Journal of Psycholinguistic Research* 27.2: 203-233. DOI: 10.1023/a:1023202116609.
- Forster, Kenneth I.; Davis, Chris. 1984. Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10.4: 680-698. DOI: 10.1037/0278-7393.10.4.680
- Friedmann, Naama; Biran, Michal; Dotan, Dror. 2013. Lexical retrieval and its breakdown in aphasia and developmental language impairment. En C. Boeckx y K. Grohmann, eds. *The Cambridge Handbook of Bilingualism*. Cambridge, R.U.: Cambridge University Press, pp. 350-374.
- Garrett, Merrill F. 1975. The analysis of sentence production. En G. H. Bower, ed. *The psychology of learning and motivation*, Vol. 9. New York: Academic Press, pp. 133-177.
- Garrod, Simon. 2006. Psycholinguistic Research Methods. En K. Brown, ed. *Encyclopedia of Language & Linguistics*. Amsterdam: Elsevier, pp. 251-257. DOI: 10.1016/B0-08-044854-2/04155-9
- Gilboy, Elizabeth; Sopena, Josep Maria. 1996. Segmentation effects in the processing of complex NPs with relative clauses. En J. E. García-Albea, *Language processing in Spanish*. Hillsdale, N.J: Lawrence Erlbaum Associates, pp. 191-206.
- Glaser, Wilhelm R. 1992. Picture naming. *Cognition* 42: 61-105. DOI: 10.1016/0010-0277(92)90040-O
- González, Julio. 1996. El fenómeno de la "Punta de la Lengua" y la recuperación léxica: estudio de sus propiedades en castellano y el efecto de la frecuencia del estímulo. *Estudios de Psicología* 56: 71-96. DOI: 10.1174/02109399660559600
- Grey, Sarah; Tagarelli, Kaitlyn M. 2018. Psycholinguistic methods. En A. Phakit, P. De Costa, L. Plonsky y S. Starfield, eds. *The palgrave handbook of Applied Linguistics research methodology*. Londres: Palgrave MacMillan, pp. 287-312. DOI: 10.1057/978-1-137-59900-1_14

- Guba, Egon G.; Lincoln, Yvonna. S. 1994. Competing paradigms in qualitative research. En N. K. Denzin y Y. S. Lincoln, eds. *Handbook of qualitative research*. Thousand Oaks, CA: Sage, pp. 105-117.
- Harley, Trevor. 2006. Speech errors: Psycholinguistic approach. En K. Brown, ed. *The Encyclopaedia of Language and Linguistics* (2ª. Ed., Vol. 11), Oxford: Elsevier, pp. 739-744.
- Heitz, Richard P. 2014. The speed-accuracy tradeoff: History, physiology, methodology and behavior. *Frontiers in neuroscience* 8: Art. 150. DOI: 10.3389/fnins.2014.00150
- Hutchison, Keith A. 2003. Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review* 10.4: 785-813. DOI: 10.3758/bf03196544
- Igoa, José Manuel. 2009. El procesamiento del léxico. En E. de Miguel, ed. *Panorama de la lexicología*. Barcelona: Ariel, pp. 405-435.
- Jegerski, Jill. 2014. Self-paced reading. En J. Jegerski y B. VanPatten, eds. *Research methods in second language psycholinguistics*. New York: Routledge, pp. 20-49.
- Jescheniak, Jörg D.; Wöhner, Stefan; Bethcke, Hanna, S.; Beaupain, Marie C. 2020. Semantic interference in the picture-word interference task: Is there a pre-lexical, conceptual contribution to the effect? *Psychonomic Bulletin & Review* 27.4: 373-378.
- Katz, Leonard; Brancazio, Larry; Irwin, Julia; Katz, Stephen; Magnuson, James; Whalen, Douglas H. 2011. What lexical decision and naming tell us about reading. *Reading and Writing* 25.6: 1259-1282. DOI: 10.1007/s11145-011-9316-9
- Lázaro, Miguel; Illera, Víctor; Sainz, Javier S. 2018. Priming effects in the recognition of simple and complex words and pseudoword. *Psicológica* 39: 198-222. DOI: 10.2478/psicolj-2018-00
- Levelt, Willem J. M.; Roelofs, Ardi; Meyer, Antje S. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22.1: 1-38. DOI: 10.1017/S0140525X99001776
- Lewis, Shevaun; Phillips, Colin. 2015. Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research* 44: 27-46. DOI: 10.1007/s10936-014-9329-z
- Macagno, Fabrizio; Bigi, Sarah. 2018. Types of dialogue and pragmatic ambiguity. En S. Oswald et al., eds. *Argumentation and language – Linguistic, cognitive and discursive explorations*. Cham, Suiza: Springer, pp. 191-218.
- Mahowald, Kyle; James, Ariel.; Futrell, Richard; Gibson, Edward. 2016. A meta-analysis of syntactic priming in language production. *Journal of Memory and Language* 91: 5-27. DOI: 10.1016/j.jml.2016.03.009
- Mathey, Stéphanie; Doignon-Camus, Nadège; Chetail, Fabienne. 2013. Syllable priming with pseudowords in the lexical decision task. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale* 3: 205-214. DOI: 10.1037/a0032456
- Meseguer, Enrique; Carreiras, Manuel. 1999. Procesamiento de ambigüedades sintácticas. En F. Cuetos y M. de Vega, eds. *Psicolingüística del español*. Madrid: Trotta, pp. 163-204.
- Mitchell, Don C.; Green, David W. 1978. The effects of context and content on immediate processing in reading. *Quarterly Journal of Experimental Psychology* 30.4: 609-636.

- Mota, Sergio; Igoa, José Manuel. 2017. Parsing complex noun phrases: Effects of hierarchical structure and sentence position on memory load. *The Spanish Journal of Psychology* 20: Article E37. DOI: 10.1017/sjp.2017.32
- Nicol, Janet; Swinney, David. 1989. The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18: 5-20. DOI: 10.1007/BF01069043
- Nicol, Janet; Swinney, David; Love, Tracy; Hald, Lea. 2006. The on-line study of sentence comprehension: An examination of dual task paradigms. *Journal of Psycholinguistic Research* 35: 215-231. DOI: 10.1007/s10936-006-9012-0
- Perea, Manuel; Gotor, Arcadio. 1997. Associative and semantic priming effects occur at very short stimulus-onset asynchronies in lexical decision and naming. *Cognition* 62.2: 3-240. DOI: 10.1016/S0010-0277(96)00782-2
- Roberts, Leah. 2014. Cross-modal priming with sentences. En J. Jegerski, y B. VanPatten, eds. *Research Methods in Second Language Psycholinguistics*. New York: Routledge, pp. 212-230.
- Rodd, Jennifer; Gaskell, Gareth; Marslen-Wilson, William. 2002. Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language* 46.2: 245-266. DOI: 10.1006/jmla.2001.2810
- Sánchez-Casas, Rosa María; Igoa, José Manuel; García-Albea, José Eugenio. 2003. On the representation of inflections and derivations: Data from Spanish. *Journal of Psycholinguistic Research* 32.6: 621-668. DOI: 10.1023/a:1026123315293
- Shao, Zeshu; Meyer, Antje S. 2017. Word priming and interference paradigms. En A.M.B. De Groot y P. Hagoort, eds. *Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide*. Oxford: Wiley-Blackwell, pp. 111-128.
- Skidelsky, Liza. 2007. La naturaleza y el contenido de los estados subpersonales de la competencia lingüística. *Subjetividad y Procesos Cognitivos* 10: 271-295.
- Soames, Scott. 1984. Linguistics and psychology. *Linguistics and Philosophy* 7: 155-179.
- Swinney, David A. 1979. Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning & Verbal Behavior* 18.6: 645-659. DOI: 10.1016/S0022-5371(79)90355-4
- Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Vigliocco, Gabriella; Butterworth, Brian; Semenza, Carlo. 1995. Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language* 3: 186-215.
- Viso, Susana del. 2002. Los lapsus linguae como fuente de datos en el estudio de la producción del lenguaje: un corpus de errores en castellano. *Anuario de Psicología* 33.3: 355-384.
- Wasow, Thomas; Perfors, Amy; Beaver, David. 2005. The puzzle of ambiguity. En O. Orgun y P. Sells, eds. *Morphology and the web of grammar: Essays in memory of Steven G. Lapointe*. Stanford, CA: CSLI Publications, pp. 265-282.
- Wickelgren, Wayne A. 1977. Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica* 41: 67-85.
- Wilson, Maximiliano A.; Cuetos, Fernando; Davies, Robert; Burani, Cristina. 2013. Revisiting Age-of-Acquisition effects in Spanish visual word recognition: The

role of item imageability. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39.6:1842-1859. DOI: 10.1037/a0033090

¹ Algunos de estos interrogantes tienen un sabor más filosófico que lingüístico como, por ejemplo, la pregunta de si el conocimiento del lenguaje tiene un carácter intencional, es decir, si posee contenidos o se refiere a algo distinto de sí mismo, y ello a pesar de su carácter inconsciente y de su naturaleza procedimental, que lo caracteriza más como un “saber cómo” que como un “saber qué” (Skidelsky 2007). Otros, en cambio, pertenecen más claramente al ámbito de la Lingüística, como la cuestión del formato que tienen las reglas que constituyen ese conocimiento, sean enunciados prescriptivos, sistemas de producción o reglas de reescritura (como los enunciados del tipo “si A, entonces B”) u operaciones de cómputo abstracto, como el mecanismo denominado *Merge* (ensamble) del programa minimista chomskiano (Eguren y Fernández Soriano 2004). Finalmente, otros parecen más dependientes de una contrastación empírica, como la posible realización neurológica de las reglas y representaciones postuladas por la Lingüística teórica (Berwick et al. 2012). Sin embargo, todos ellos comparten un problema metateórico, que estriba en decidir cuáles de estas cuestiones están sujetas a demostración o refutación empírica, cuáles dependen de una argumentación teórica y cuáles son axiomáticas, es decir, teóricamente necesarias y previas a cualquier argumentación teórica o confirmación empírica.

² Las oraciones de vía muerta (*garden-path*, en inglés) son oraciones que generan una decisión estructural que resulta ser incorrecta y que obliga a reanalizar la estructura y el significado de la oración que se está procesando. Un ejemplo de este tipo de oraciones en castellano es “La cliente le pedirá al camarero *que le sirva la comida* que se dé prisa”, en la que la cláusula *que le sirva la comida*, inicialmente etiquetada como cláusula completiva, debe ser reanalizada como cláusula de relativo al recibir el fragmento final de la oración, que es el verdadero complemento del verbo *pedirá*.

³ Las oraciones con inclusión o incrustación central (*center-embedded sentences*, en inglés) son aquellas que introducen constituyentes subordinados (p.ej. cláusulas de relativo) en el interior de otros constituyentes, por ejemplo, entre el sujeto y el verbo principal de la oración, creando estructuras anidadas, como en la oración “El ratón [que el gato [que el perro perseguía] acechaba] consiguió escapar” (los corchetes indican los límites de los constituyentes anidados). Este proceso provoca una sobrecarga de memoria que impide o dificulta el análisis y la comprensión de la oración. Si los constituyentes se disponen con ramificación hacia la derecha, la oración se puede procesar con mayor facilidad; así “El perro perseguía al gato [que acechaba al ratón [que consiguió escapar]]”.

⁴ Cabe mencionar que también hay una versión auditiva de la tarea de decisión léxica, en la que se emplean palabras y pseudopalabras presentadas en forma oral (Ernestus y Cutler 2015), si bien esta modalidad auditiva de presentación se emplea con menor frecuencia que la escrita.

⁵ La palabra inglesa *priming* no tiene una traducción definida en castellano. A veces se traduce como “facilitación”, dado que a menudo facilita la respuesta del sujeto reduciendo su latencia o favoreciendo unas respuestas frente a otras, pero muchas veces provoca efectos inhibitorios, lo que invalida esa traducción. Otras veces se traduce por “imprimación” o “primado”, un neologismo que conserva una semejanza formal con el original en inglés acoplado a la morfología del castellano. Aquí optamos por mantener el término original por ser el que se usa con mayor frecuencia en los textos en castellano.

⁶ Existe una herramienta en la red, denominada *Wuggy*, diseñada para generar pseudopalabras en diversos idiomas –vasco, holandés, inglés, francés, alemán, serbio y croata en los alfabetos cirílico y latino, respectivamente, castellano y vietnamita– con el propósito de emplearlas en experimentos de Psicolingüística. Este es el enlace a la página web: <http://crr.ugent.be/programs-data/wuggy>

⁷ Un “tonema” se define como “la inflexión que recibe la unidad melódica a partir de la última sílaba acentuada” (Diccionario de la lengua española, RAE: <https://dle.rae.es/>).

Aportación de las investigaciones con neuroimagen funcional a los estudios lingüísticos: algunas preguntas y respuestas

Victoria Marrero-Aguar
UNED
vmarrero@flog.uned.es

Resumen

El objetivo principal de este artículo, de carácter meramente introductorio, es mostrar el interés lingüístico de resultados obtenidos mediante las principales técnicas de neuroimagen funcional aplicadas al estudio del lenguaje, y los retos asociados al encuentro interdisciplinar entre la Lingüística y la Neurociencia. Comenzamos resumiendo los principios básicos de funcionamiento de las dos metodologías más utilizadas actualmente para estudiar la actividad del cerebro mientras realiza tareas lingüísticas: la de base electromagnética y la de base metabólica. A continuación, propondremos ejemplos de preguntas relevantes para la Lingüística que pueden encontrar respuestas en datos obtenidos mediante técnicas de neuroimagen. Concluiremos reflexionando sobre los distintos niveles de implicación de los lingüistas en el uso y avance de las técnicas de neuroimagen.

Palabras clave: Neurolingüística, Biolingüística, técnicas de neuroimagen funcional, procesamiento cerebral del lenguaje.

Abstract

The main objective of this article, of a merely introductory nature, is to show the linguistic interest of the results obtained by means of the main functional neuroimaging techniques applied to the study of language, and the challenges associated with the interdisciplinary encounter between Linguistics and Neuroscience. We begin by summarising the basic principles of the two most widely used methodologies to study the functioning of the brain while performing linguistic tasks: the electromagnetic-based and the metabolic-based techniques. We will then propose examples of questions relevant to linguistics that can be answered by data obtained using neuroimaging techniques. We will conclude by reflecting on the different levels of involvement of linguists in the use and progress of neuroimaging techniques.

Key words: Neurolinguistics, Biolinguistics, functional neuroimaging techniques, brain language processing.

1. Introducción

La historia de las relaciones entre Lingüística y Neurociencia empieza en abril de 1861, en una mesa de autopsias. Pocos días antes había fallecido monsieur Leborgne, un hombre de 51 años que perdió el habla a los 30, tras una lesión en mitad del lóbulo frontal del hemisferio izquierdo, y que vivió las dos décadas siguientes sin emitir más que la sílaba /tan¹. Cuando su cerebro fue presentado por el padre de la Neurociencia, Paul Broca, ante

la Societé d'Antropologie de París (fundada por él mismo, Giménez-Roldán 2017), comenzó una larga y fructífera línea de investigación sobre las bases neuroanatómicas de la capacidad humana del lenguaje.

Pero tuvieron que pasar casi 70 años antes de que las investigaciones pudieran realizarse con personas vivas y, por tanto, capaces de hablar mientras estaban siendo estudiadas. Fue en abril de 1929 cuando Hans Berger publicó el resultado de unos registros que habían comenzado cinco años antes, colocando un osciloscopio de dos canales sobre la piel que cubría una trepanación craneal en un joven de 17 años (La Vaque, 1999): entonces comenzó la electroencefalografía (EEG) como técnica de estudio del cerebro vivo. Cincuenta años más tarde, en 1975, se presenta la Tomografía por Emisión de Positrones (PET, por sus siglas en inglés), y en 1979 se graban las primeras imágenes por resonancia magnética. Ya en la década de los 90 se multiplican las opciones, al sumarse las imágenes obtenidas por resonancia magnética funcional (fMRI)² y mediante técnicas ópticas³. Desde hace casi un siglo, por lo tanto, el estudio de la actividad cerebral en humanos vivos se ha convertido en una de las áreas más productivas de la ciencia. Actualmente contamos con métodos y técnicas no invasivas que permiten observar la respuesta neuronal ante tareas de todo tipo, y entre ellas, la actividad lingüística ocupa un lugar destacado, tanto en su dimensión oral como lectoescrita. Como botón de muestra de ese interés creciente entre las relaciones del lenguaje con las técnicas de neuroimagen, las referencias que arroja la búsqueda de los términos “language AND neuroimaging” en la base de datos PsycINFO se ha multiplicado por más de 100 en las tres últimas décadas: de 13 publicaciones en el quinquenio 1988-1992 pasamos a 1413 entre 2013 y 2018. Los resultados de la búsqueda en la base de datos Scopus de “Language” y “Neuroimaging” en el título, el resumen o las palabras clave de los trabajos se recogen en la Figura 1.

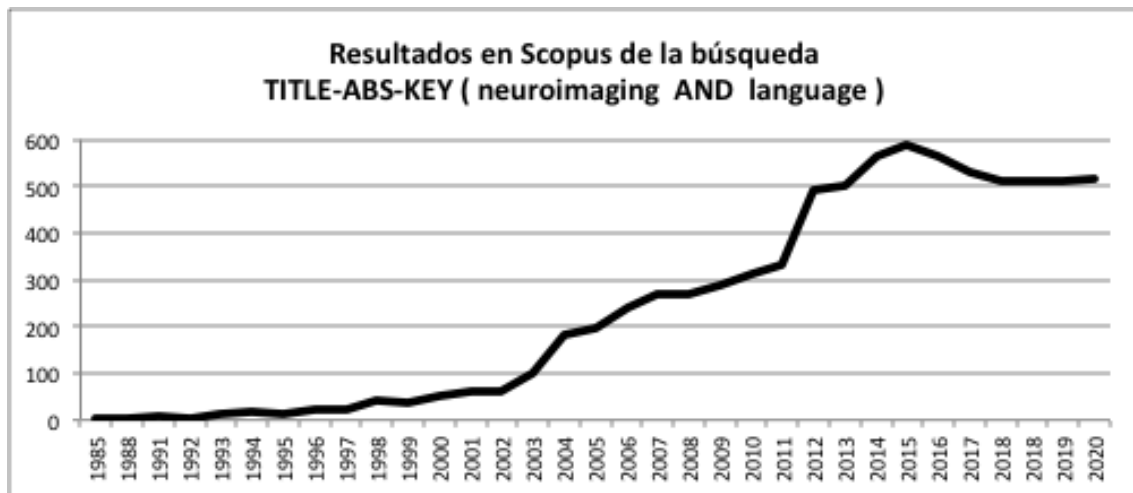


Figura 1. Resultados en la base de datos Scopus de las palabras clave “neuroimaging and language” cuando aparecen en el título, el resumen o las palabras clave del trabajo. En el eje horizontal se muestran los años de publicación y en el vertical, el número de publicaciones

En definitiva, el lenguaje, como actividad humana por excelencia, es uno de los objetos de estudio más relevantes para la Neurociencia cognitiva actual; los resultados que nos proporcionan las técnicas de neuroimagen, aunque generalmente se enmarcan en experimentos no diseñados por lingüistas, arrojan datos de gran interés para la Lingüística, y son objeto de mejora y revisión constantes, así como un buen punto de partida para un análisis crítico de las antiguas certezas (por ejemplo, Tremblay y Dick

2016). No solo la Neurolingüística, sino también la Biolingüística o la Psicolingüística, disciplinas cercanas en el ámbito cognitivo, llevan décadas haciendo un uso pleno de estas técnicas, y contribuyendo a su desarrollo (Paredes Duarte y Varo Varo 2006). Incluso áreas cuyo foco no es el lenguaje como capacidad cognitiva, sino la relación entre lengua y sociedad, como la etnolingüística o la sociolingüística, también están creciendo en su intersección con la Neurociencia⁴. En la próxima sección analizaremos cómo los niveles de análisis del lenguaje en torno a los cuales se estructuran las áreas centrales de nuestra disciplina (la fonología, la gramática, la semántica, la teoría lingüística...) también pueden encontrar nuevas respuestas a sus interrogantes clásicos en los estudios con técnicas de neuroimagen.

2. Principales técnicas para el estudio del cerebro ante el lenguaje

Esta sección pretende facilitar un primer acercamiento a la lectura crítica de la bibliografía procedente de estudios con técnicas de neuroimagen. Para complementar esta sección con recursos audiovisuales, recomendamos la charla TED2014 de Nancy Kanwisher *A neural portrait of the human mind*, en la que se muestran numerosas imágenes que pueden facilitar la comprensión de algunos aspectos técnicos.

Este apartado está estructurado cronológicamente: comenzaremos con las técnicas de base electromagnética, que son las más antiguas (pero siguen plenamente vigentes) y finalizaremos con las de base metabólica, entre las que se encuentran los últimos desarrollos aplicados al estudio del lenguaje y demás procesos cognitivos.

2.1. Técnicas de base electromagnética

La actividad de las neuronas es eléctrica por naturaleza. Las técnicas de base electromagnética miden de forma directa esa actividad eléctrica o el campo magnético asociado a la misma. Comenzaremos por las que detectan la electricidad, que proporcionan datos interesantes sobre el procesamiento del lenguaje a lo largo del tiempo con precisión de milisegundos, pero no permiten apreciar qué áreas se han activado para ello; terminaremos con la que detecta el campo magnético, que sí permite identificar grandes áreas del cerebro activadas, manteniendo al tiempo una resolución temporal muy buena.

El **electroencefalograma (EEG)** permite registrar, gracias a unos electrodos inocuos que se adhieren al cuero cabelludo, las ondas cerebrales que atraviesan el cráneo, resultado de la actividad conjunta de la corteza (o de un hemisferio, generalmente) en frecuencias que van desde unos 2 - 4 Hz en las ondas delta hasta más de 30 Hz en las gamma, pasando por las theta, alfa y beta. En los últimos años el estudio de estas ondas (estructurado en lo que se conoce como la perspectiva oscilómica) ha desarrollado propuestas muy interesantes sobre sus relaciones con el procesamiento del lenguaje: la banda delta estaría sincronizada con la información prosódica, la banda theta con la silábica y la banda gamma con la fonológica, y los acoplamientos en fase de las bandas delta y theta estarían relacionados con el procesamiento semántico y sintáctico (Murphy 2018). Puede verse una revisión excelente sobre esta perspectiva en Benítez-Burraco y Murphy (2019), y profundizar en Murphy (2020).

Por otra parte, son innumerables los trabajos realizados midiendo el comportamiento de unas ondas eléctricas concretas en el momento en que se presentan los estímulos lingüísticos: se trata de los potenciales evocados, y más concretamente los relacionados con eventos, ERP, (*Event Related Potentials*). Algunos de estos picos de energía son el N400, que se relaciona con la integración semántica; el P600, que al igual que el LAN (*left anterior negativity*) y el ELAN (*early left anterior negativity*), aparece ante errores, ambigüedades o incoherencias sintácticos, y el MMN (*mismatch negativity*), que se da cuando se nos presenta un estímulo auditivo diferente de los anteriores. Puede verse una revisión sobre el papel de los ERP en estudios lingüísticos en Beres (2017)⁵. Por último, mencionaremos brevemente la estimulación directa transcraneal (tDCs), una técnica se emplea con frecuencia en el tratamiento logopédico de la afasia, o en la depresión y otros trastornos psiquiátricos (respectivos metaanálisis en Corrales-Quispircra et al. 2020; Palm et al. 2016; Kekic et al. 2016); con población sana se ha utilizado en tareas lingüísticas de denominación, fluidez verbal, lectura, etc. (pueden consultarse más detalles en la revisión de Monti et al. 2013), aunque desde hace unos años su efectividad se ha puesto en duda (Westwood 2017).

No entraremos en más detalles sobre estos instrumentos por dos razones: la primera es que este monográfico cuenta con un artículo completo dedicado a la EEG y, en concreto, a los ERPs (véase Haro 2022); la segunda es porque no son, en realidad, técnicas de neuroimagen, ya que no permiten deducir qué áreas del cerebro son las que se han activado. Pese a ello, la EEG es el sistema más antiguo y posiblemente el más utilizado en Neurolingüística, así como el más accesible económicamente y el que se instala más fácilmente. Además, lo que pierde en resolución espacial lo compensa en resolución temporal, puesto que detecta los cambios en los potenciales con una sensibilidad de milésimas de segundo.

A diferencia de la EEG, la **electrocorticografía (EcoG)** capta mejor las ondas eléctricas neuronales y permite asignar la actividad a zonas concretas del cerebro porque suprime la barrera del cráneo, al colocarse los electrodos directamente en la corteza cerebral. Esta herramienta, unida a una interfaz computacional (*Brain-Computer Interface, BCI system*) que se ha desarrollado muy recientemente (el primer caso fue presentado por Vansteensel et al. en 2016), permite transformar la actividad registrada en habla, convirtiendo en fonemas ciertos patrones de actividad en la superficie cortical (Ramsey et al. 2018). Sin embargo, para utilizarla es necesaria una trepanación craneal, una intervención quirúrgica muy invasiva que hace este método útil solo para personas cuya patología lo justifique.

La **magnetoencefalografía (MEG)**, por su parte, registra los campos magnéticos que acompañan a la actividad eléctrica de las neuronas, tras su paso por el cráneo y el cuero cabelludo. A esas medidas directas se les aplican procedimientos de reconstrucción de la fuente que permiten una representación anatómica de las áreas de la corteza que se han activado, de modo que sí forma parte de las técnicas de neuroimagen clásicas, aunque su resolución espacial es discreta. La ventaja de los campos magnéticos, frente a la actividad eléctrica, es que no sufre distorsiones al atravesar los tejidos biológicos, pero su principal limitación es la bajísima intensidad del campo magnético del cerebro, de 10^8 o 10^9 menor que la del campo magnético terrestre o la de un frigorífico casero. De hecho, solo las dendritas apicales de las células piramidales consiguen ser registradas por los numerosos sensores del casco de un magnetoencefalógrafo. Aunque este aparato es mucho más costoso que el EEG (en todos los sentidos, económico, tecnológico y por dificultad de uso) y requiere que el participante guarde una inmovilidad casi absoluta (hasta un

pestañeo puede generar interferencias), ofrece una excelente resolución temporal, y cierta información espacial. Por ello se ha convertido en otra de las habituales fuentes de información para estudios sobre el lenguaje, como muestra la revisión de Frye et al. (2009) o la más reciente de Dikker et al. (2020).

Por último, la **estimulación magnética transcraneal (TMS)** también emplea el electromagnetismo para investigar el cerebro. No obstante, esta técnica no se limita a recoger los cambios en la actividad de las neuronas tras la presentación de un estímulo, sino que genera pulsos magnéticos que se aplican sobre la cabeza del participante, lo cual desencadena una breve corriente eléctrica en la superficie de la corteza capaz de despolarizar la población de neuronas cercanas. La TMS se utiliza especialmente para comprobar, antes de intervenciones quirúrgicas indicadas en epilepsias graves o en tumores cerebrales, las repercusiones funcionales (lingüísticas o de otro tipo) del área que se va a someter a cirugía. Puede verse una revisión sobre su uso en estudios lingüísticos en Miniussi et al. (2010).

2.2. Técnicas de base metabólica

La otra posibilidad para descubrir qué zonas del cerebro se activan en el momento de realizar una determinada tarea es averiguar dónde es mayor el consumo metabólico. En este caso, se trata de medidas indirectas basadas en la premisa de que la actividad neuronal desencadenada por los procesos cognitivos incrementa el consumo de oxígeno y glucosa en las zonas activas del cerebro (es lo que se conoce como *acoplamiento neurovascular*). Los cambios en el nivel de oxígeno en sangre provocan variaciones en el magnetismo de la hemoglobina que son captadas en una resonancia magnética; y también diferencias en la absorción o reflexión de rayos de luz infrarroja, detectadas a su vez por los métodos ópticos. Por otra parte, para observar el consumo de glucosa es necesario inyectar al sujeto un *trazador*, esto es, una sustancia radiactiva capaz de visibilizar esa actividad, como se detallará a continuación. Todas estas técnicas ofrecen una buena resolución espacial, pero peor resolución temporal que las derivadas de la actividad eléctrica, ya que dependen de procesos metabólicos desencadenados tras la activación neuronal.

La primera gran técnica tradicional entre las de base metabólica es la ya mencionada **tomografía por emisión de positrones (PET)**. Requiere que el sujeto ingiera, inhale o se inyecte un trazador radiactivo, cuya función es unirse a la glucosa para hacerla detectable (en concreto, aporta los positrones que se aniquilarán al entrar en contacto con los electrones de la corteza cerebral, generando los fotones que capta el tomógrafo, la máquina que permite obtener las PET). Aunque los trazadores tienen una vida muy corta⁶, no dejan de presentar cierto carácter invasivo; además, para ser producidos requieren un acelerador de partículas (un ciclotrón), lo cual, unido al escáner, encarece enormemente la instalación. A pesar de ello, se ha utilizado en estudios lingüísticos muy variados y, aprovechando que es silencioso, especialmente en el ámbito auditivo, en relación con la percepción de los sonidos del habla y el procesamiento fonológico (Johnsrude et al. 2002). Para una presentación general de las investigaciones sobre el lenguaje que utilizan esta técnica puede leerse el capítulo de Horwitz y Wise (2008) en el *Handbook of the Neuroscience of Language*.

Sin embargo, la técnica estrella en neuroimagen del lenguaje es, posiblemente, la imagen por **resonancia magnética funcional (fMRI)**⁷. Para obtener representaciones gráficas utilizando el fMRI es necesario, en primer lugar, generar mediante grandes imanes un

potentísimo campo magnético⁸ capaz de polarizar los átomos de hidrógeno del cuerpo (y del cerebro). Además, se necesita un sistema de radiofrecuencia para enviar una onda de radio (128 MHz para 3T) que los desestabilice. La máquina detecta el momento en que estos átomos vuelven a alinearse con el campo magnético (el proceso es un poco complicado, pero Armony et al. 2012 lo explican de forma muy accesible). En el ámbito médico, y con fines diagnósticos, se utiliza generalmente la resonancia magnética estática; aunque la que nos interesa a los lingüistas es la calificada como *funcional*, porque se registra mientras el sujeto realiza alguna tarea, y mide las diferencias magnéticas entre la sangre oxigenada empleada por las neuronas activas y la desoxigenada de las inactivas. Una de las principales limitaciones de la fMRI es que la máquina genera un ruido tremendo (de al menos 90-110 Hz), que hace imprescindible la protección auditiva y dificulta la presentación de estímulos sonoros. Aunque se están haciendo muchos esfuerzos para solventar esta limitación (Ljungberg et al. 2021), y desde hace años se han publicado trabajos con los estímulos presentados auditivamente (por ejemplo, Holmes et al. 2021), lo cierto es que en la bibliografía sobre procesamiento lingüístico predominan las tareas que requieren lectura (sobre cuyas implicaciones volveremos en las conclusiones).

La resonancia magnética también está en la base de unas técnicas que permiten visibilizar no ya la actividad neuronal, sino la de la sustancia blanca del cerebro, la que compone las fibras mielinizadas de los axones que conectan las neuronas. Se pueden obtener así representaciones de las conexiones neuronales, de las redes que unen diferentes áreas del cerebro: son las imágenes de **resonancia magnética por difusión (dMRI)** y las **imágenes con tensor de difusión (DTI)**, que constituyen las técnicas de **tractografía**. Tanto la dMRI como la DTI captan la disposición de los protones del agua en el cerebro. Los axones de las neuronas están recubiertos por una capa aislante, la mielina, que limita las posibilidades de difusión del agua en su interior, y por tanto influye en el alineamiento de esos protones. Los cambios de dirección en los protones son detectados por el escáner, y tras aplicar una compleja computación acaban representándose como líneas de colores. La tractografía está en la base del *Human Connectome Project*, un gran reto internacional en Neurociencia, abordado entre 2009 y 2014, cuyo objetivo fue crear y compartir la “matriz de conexiones” del cerebro humano. Desde 2015 el proyecto, que se puede consultar en la web <https://www.humanconnectome.org>, ha continuado creciendo en dos áreas: el *conectoma a lo largo de la vida (lifespan connectome)* con datos de casi cinco mil sujetos sanos de todas las edades, y los *conectomas relacionados con la enfermedad humana* (Alzheimer, demencia, psicosis, epilepsia, pérdida de visión, etc.). Uno de sus resultados es el *conectoma del lenguaje* (Dick et al. 2014; Vassal et al. 2016; Przeździk et al. 2019), que se resume en esta cita:

El lenguaje se procesa en un sistema cortical y subcortical distribuido. Aunque es importante identificar las regiones cerebrales que forman parte de este sistema, es igualmente importante establecer la conectividad anatómica que sustenta sus interacciones funcionales. El marco más prometedor es aquel en el que el lenguaje se procesa a través de dos “corrientes” que interactúan —una corriente dorsal y otra ventral— ancladas mediante largas vías de fibras de asociación (Dick et al. 2014: 455; esta traducción y todas las que siguen son propias).

El conectoma del lenguaje se ha relacionado con el procesamiento en personas bilingües (Mitsuhashi et al. 2020), otra de las áreas de aplicación de las técnicas tractográficas

(Mohades et al. 2012; García-Pentón et al. 2014 con hablantes de español y euskera; Hämäläinen et al. 2017; Pliatsikas et al. 2020; etc.)

El último grupo de técnicas de neuroimagen funcional basadas en los cambios metabólicos tiene una base óptica: son la **espectroscopía funcional del infrarrojo cercano (fNIRS)** y la **tomografía óptica difusa de alta densidad (HD-DOT)**. Ambas miden las variaciones en la absorción o reflexión de rayos de luz infrarroja producidas por los cambios en el nivel de oxígeno en la sangre. En sus revisiones sobre las aplicaciones de estas técnicas a los estudios lingüísticos, tanto Peelle (2017) como Quaresima et al. (2012) o Scherer et al. (2009, en español) destacan sus principales ventajas: aunque no sobresalgan ni por su resolución espacial (mejor en HD-DOT que en fNIRS) ni por la temporal, son equipamientos relativamente asequibles, silenciosos y con cierto grado de portabilidad que permite a los participantes algún movimiento mientras realizan la tarea e, incluso, la interacción con otras personas. Por eso se han utilizado para investigar el desarrollo del lenguaje en bebés y niños, en poblaciones que no pueden ser sometidas a imágenes por resonancia magnética, como los implantados cocleares, y en estudios de comprensión de habla sin interferencias de ruido.

En suma, cada técnica tiene sus ventajas y sus inconvenientes, y es bastante frecuente el uso combinado de algunas de ellas. Así, encontramos centenares de trabajos en cuyo título aparecen simultáneamente la MEG y el EEG, y en manuales como el de Hari y Puce (2017) se explica su uso conjunto. No pocos añaden además la fMRI (entre otros, Wang et al. 2021), cuya combinación con la PET es también bastante habitual (Sekiyama et al. 2003, por ejemplo). Horwitz y Poeppel (2002) proponen incluso la utilización conjunta de estas cuatro técnicas. Entre las de tipo óptico, las combinaciones más frecuentes son las de fNIRS y fMRI (como muestra, Yuan y Ye 2013). La finalidad es aprovechar los puntos fuertes de cada una. En la Figura 2 (adaptada de Lizarazu 2017) se comparan la resolución temporal y espacial de cinco de estos métodos: los de base eléctrica comparten su excelente comportamiento en el eje del tiempo, con diferencias en el del espacio (peor el EEG, regular la MEG y buena la EcoG); las de base metabólica tienen todas peor resolución temporal (la peor, la PET, seguida por la fMRI, y la NIRS), pero mejor resolución espacial (especialmente las dos primeras). Por último, la Figura 3 (de elaboración propia y, en cierta medida, subjetiva) compara gráficamente el coste económico de cada equipamiento (calculado *grosso modo*) y el nivel de invasividad o incomodidad (ruido, inmovilidad...) que provoca en los participantes: en un solo vistazo es fácil apreciar las ventajas de la fNIRS o las metodologías basadas en el EEG.

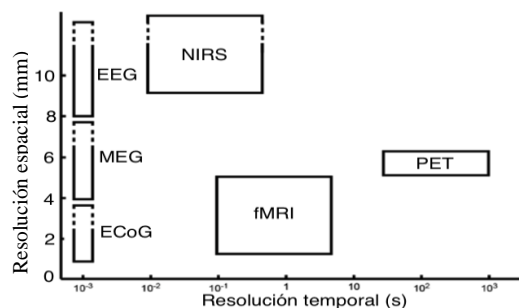


Figura 2. Resolución espacial y temporal de diferentes técnicas de neuroimagen. Adaptado de Lizarazu (2017, Figura 9)

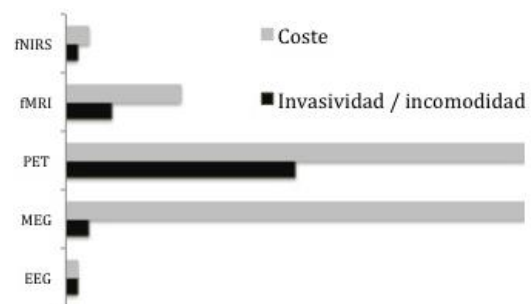


Figura 3. Coste relativo aproximado y nivel de invasividad o incomodidad de diferentes técnicas de neuroimagen

3. ¿Qué preguntas relevantes para la Lingüística pueden tener respuesta en estudios con neuroimagen? Algunos botones de muestra

Tras la somera presentación de las técnicas de neuroimagen utilizadas en la actualidad para el estudio de la cognición humana en general, y el lenguaje en particular, en este apartado ejemplificaremos con algunas preguntas concretas de investigación qué cuestiones lingüísticas pueden recibir nueva luz a partir de los datos proporcionados por estos instrumentos. Por lo tanto, presentaremos unos pocos ejemplos que solo pretenden poner de manifiesto qué tipo de evidencias nos pueden proporcionar los estudios con neuroimagen en relación a objetos de estudio habituales en Lingüística.

Utilizaremos aquí una progresión clásica: comenzaremos en el nivel fónico, luego abordaremos el morfológico, el sintáctico y terminaremos con el semántico. Todo ello sin la menor pretensión de exhaustividad; para conocer en mayor profundidad las bases neuroanatomofisiológicas del procesamiento lingüístico remitimos, en español, a Benítez-Burraco (2006), o, para una perspectiva estrictamente médica, a González y Hornauer-Hughes (2014).

3.1. En el nivel fónico: ¿cómo pasamos de la fonética a la fonología? ¿Podemos definir los fonemas solo con rasgos acústicos, o son imprescindibles también los rasgos articulatorios?

¿En qué momento y de qué manera conseguimos que la onda sonora que llega a nuestro oído se convierta en una sucesión de fonemas, unidades discretas y abstractas, con capacidad para generar pares mínimos en las lenguas? Encontramos esta pregunta al menos desde los inicios de la Lingüística moderna. En el *Cours de linguistique générale* de Ferdinand de Saussure se alude a una distinción entre la sensación acústica y su representación mental: el sonido solo se incorpora a la estructura lingüística cuando se ha transformado en un elemento psíquico, la imagen verbal (Saussure 1916: 28-29).

La respuesta que nos ofrece la neuroimagen funcional permite una reflexión abstracta sobre la existencia del fonema como unidad lingüística (Rampinini y Ricciardi 2017), pero también descubrir dónde tiene lugar esa mutación de señal continua en elementos discretos: en la corteza auditiva primaria (*primary auditive cortex* o PAC, correspondiente a las áreas de Broadman 41 y 42), que en humanos diestros suele estar más desarrollada en el hemisferio izquierdo, aunque en los primates es simétrica en ambos hemisferios. Y lo que es más interesante, permite mostrar cómo se produce esta mutación: gracias a la organización tonotópica de esa zona del cerebro, también presente en la cóclea y en las vías auditivas. Las neuronas en la PAC están especializadas para responder a determinadas frecuencias de los sonidos, se distribuyen en un mapa ordenado, desde los tonos más bajos, en un extremo, a los más altos, en el extremo contrario, y se comportan de forma distinta en el hemisferio izquierdo que en el derecho.

[la secuencia de] tonos de altos a bajos se organiza de manera medial a lateral y rostral a posterior en ambas PAC. La PAC izquierda parece ser más sensible a los tonos entre 500 y 4000 Hz que la derecha; la derecha es, en general, más sensible a las frecuencias más altas [...]. Cabe destacar que los fonemas, los sonidos del lenguaje, se encuentran principalmente en el rango de frecuencias de 500 a 4000 Hz (Bernal y Ardila 2016: 458).

La corteza auditiva primaria del hemisferio izquierdo, especializada en los sonidos del habla, tiene capacidad para inhibir la señal procedente del hemisferio derecho y además recibe información directa de la corteza visual, permitiendo una descodificación multimodal de la señal: “para la percepción categorial de los fonemas, tanto auditiva como visualmente, la PAC está fuertemente lateralizada en el lóbulo temporal izquierdo” (Bernal y Ardila 2016: 459). La naturaleza intrínseca del fonema, pues, sería no solo auditiva sino multimodal.

La segunda pregunta que planteamos en este epígrafe, esto es, cuál es el papel de los procesos articulatorios en la definición de los fonemas, se remonta igualmente a las propuestas del maestro ginebrino: “no podemos, pues, reducir la lengua al sonido ni separar el sonido de la articulación bucal” (Saussure 1916: 24). La Lingüística posterior continuó considerando la cuestión en infinidad de ocasiones, desde la primera fonología generativista, cuando Chomsky y Halle (1968) incorporan rasgos distintivos de naturaleza articulatoria a los puramente acústicos de Jakobson et al. (1952), hasta la Teoría Motora de la percepción del habla, cuya primera hipótesis es que los objetos de la percepción del habla no son los sonidos, sino los gestos articulatorios necesarios para generarlos (Lieberman y Mattingly 1985; Galantucci et al. 2006). El descubrimiento de las neuronas espejo, que se activan no solo para ejecutar ciertas acciones, sino también cuando se ven o se oyen esas acciones ejecutadas por otros (Rizzolatti et al. 1996) se relacionó desde muy pronto con la Teoría Motora. En este sentido, varios estudios con técnicas de neuroimagen han mostrado que las áreas motoras del cerebro necesarias para la producción de los sonidos están activas durante la percepción del habla (Watkins et al. 2003; Wilson et al. 2004; Wilson y Iacoboni 2006) o que su estimulación interfiere con la descodificación del mensaje, especialmente en condiciones difíciles, como en presencia de ruido o distorsiones (Rogers et al. 2014; Schomers y Pulvermüller 2016; Nuttall et al. 2018). Sin embargo, también encontramos evidencias en sentido contrario: personas con lesiones en las áreas atribuidas a las neuronas espejo no tienen problemas en la percepción del habla (Rogalski et al. 2011; Stasenko et al. 2013). En definitiva, la cuestión sigue abierta, tanto en Lingüística como en Neurociencia, y ambas áreas se enriquecen la una a la otra con sus respectivos avances.

No podemos extendernos más en el nivel fónico, pero los estudios con técnicas de neuroimagen también han arrojado luz sobre otras muchas cuestiones, algunas de alcance geolectal, como el procesamiento de diferentes variedades de una lengua⁹, incluidas el castellano y el español de América (Hernández et al. 2019). En este último trabajo se detecta que las actitudes lingüísticas (la diferencia entre la valoración de la variedad geolectal propia frente a la variedad ajena) tienen correlatos neuronales: así, los centros de recompensa en el tálamo y el centro estriado estarían más activos ante el estímulo correspondiente a la variedad propia, lo cual “presumiblemente revela el sentido de pertenencia social de las personas” (Hernández et al. 2019: 12). También se han investigado las diferencias entre la prosodia lingüística y la emocional (puede leerse un metaanálisis de 29 trabajos en Belyk y Brown 2014), e, incluso la asociación de algunos sonidos con determinados rasgos semánticos (por ejemplo, la secuencia /fl/ con ‘fluido’ en Mompeán et al. 2020): el fonosimbolismo (Kanero et al. 2014; Revill et al. 2014).

3.2. En el nivel morfológico: ¿nos cuesta lo mismo asignar el género a la palabra fuente que a carro? ¿Procesamos igual el género que el número?

La mayor riqueza morfológica de las lenguas romances respecto al inglés ha permitido que este nivel de análisis en español haya recibido una cierta atención en Neurociencia:

La investigación neurolingüística sobre las lenguas romances, principalmente el francés, el italiano y el español, ha resultado clave para el progreso de esta área, especialmente con referencias específicas a cómo funciona la infraestructura neuronal del lenguaje en el caso de sistemas flexivos más ricos que el inglés. Entre los temas de investigación más habituales se encuentran los fenómenos de concordancia; los estudios sobre el español y el italiano demostraron que la concordancia entre diferentes morfemas [*across features and domains*] (por ejemplo, la concordancia de número o de género) involucra sustratos neuronales parcialmente diferentes (Bambini y Canal 2021: 1).

La primera pregunta que planteamos en este apartado tuvo respuesta en el estudio de Hernández et al. (2004), mediante un experimento con fMRI. En esta investigación se analizó la actividad neuronal de 9 hispanohablantes ante la tarea de decidir si 96 nombres eran masculinos o femeninos: la mitad eran opacos en cuanto al género, puesto que no terminaban ni en *-o* ni en *-a*. Las respuestas ante estos estímulos requirieron mayor activación y en más áreas que las de estímulos masculinos terminados en *-o* y femeninos en *-a*. Por lo tanto, si consideramos que una activación mayor y en más áreas se relaciona generalmente con una mayor carga cognitiva, la respuesta a nuestra pregunta sería negativa: nos costaría más decidir que la palabra *fuate* es femenino que marcar *carro* como masculino.

El resultado anterior no es sorprendente, probablemente habríamos apostado por ello. Pero la respuesta a la segunda pregunta, “¿procesamos igual el género que el número?”, quizá no resulte tan evidente: detectar los errores de concordancia en número requiere activar zonas del cerebro dedicadas al procesamiento de la cantidad, algo que no ocurre con el género (Carreiras et al. 2010). Además, un estudio con magnetoencefalografía permitió a Molinaro et al. (2013) descubrir que la diferencia de significado entre singular y plural (‘uno’ / ‘más de uno’) provocaba la aparición de un componente tardío que suele reflejar incrementos en la dificultad del procesamiento semántico. Así pues, la respuesta en este caso también parece ser negativa: los errores de concordancia en ambos morfemas se procesan de forma diferente y presentan distintos niveles de dificultad.

Otra área de la morfología que ha recibido notable atención en estudios con técnicas de neuroimagen es la diferencia entre verbos y nombres (que ya había llamado la atención de los afasiólogos mucho antes: Damasio y Tranel 1993). Posteriormente se han reinterpretado las diferencias de procesamiento encontradas entre ambas clases de palabras considerando las propiedades semánticas de los diferentes grupos de verbos. Por ejemplo, en el marco de la cognición corporeizada (Barsalou 2010, véanse más referencias en el próximo apartado), el estudio de los verbos de acción ha dado lugar a la hipótesis de que procesar conceptos asociados al movimiento provoca la activación de las áreas motoras frontales, las mismas que se pondrían en funcionamiento para generar ese movimiento. En una interesante revisión crítica, Vigliocco et al. (2011) llegan a esta conclusión:

Si consideramos la confusión que hay en la mayoría de los estudios entre las distinciones semánticas (objetos frente a acciones) y la distinción gramatical (sustantivos frente a verbos), y la confusión entre los estudios relacionados con los mecanismos de procesamiento de una sola palabra y los estudios relacionados con la integración de la oración, la imagen resultante es relativamente clara: se observa una clara separabilidad neuronal entre el procesamiento de las palabras

de objeto (sustantivos) y las palabras de acción (típicamente verbos); los efectos de la categoría gramatical aparecen o se hacen más fuertes cuando las tareas y las lenguas imponen mayores demandas de procesamiento (Vigliocco et al. 2011: 407).

Concluimos esta breve presentación de las respuestas que nos ofrecen las técnicas de neuroimagen sobre el nivel morfológico con el deseo que cierra la reflexión de Kemmerer (2014) sobre la intersección entre la Lingüística y la Neurociencia cognitiva, en el sentido de una mayor atención por parte de la Neurociencia a las aportaciones procedentes de la Lingüística (concretadas, en este caso, en la tipología): “A medida que pase el tiempo, los neurocientíficos cognitivos podrían empezar a interesarse más por dilucidar los fundamentos corticales de los tipos de categorías que suelen estudiar los tipólogos, que, al fin y al cabo, son las máximas autoridades en materia de diversidad entre lenguas” (Kemmerer 2014: 43).

3.3. En el nivel sintáctico: ¿las oraciones subordinadas de relativo de sujeto y las de objeto son similares cognitivamente? ¿Es más fácil procesar *Ana apareció o Ana cortó el pan?* ¿Y *Ana cortó el pan* frente a *Ana abrió la puerta?*

Desde comienzos del siglo XXI se han publicado diversas revisiones que relacionan la gramática generativa (Chomsky 1957, 1995) con los descubrimientos proporcionados por las técnicas de neuroimagen. Posiblemente el trabajo más conocido en esta línea es el de Friederici et al. (2017), con el propio Chomsky como segundo firmante, que fue precedido por Grodzinski y Friederici (2006), Ohta et al. (2013), o por la revisión crítica de Caplan (2001). Pero en esta sección nos plantearemos preguntas muy básicas e independientes de modelos teóricos. La primera es un tema estrella en Neurolingüística: la diferencia entre dos tipos de oraciones subordinadas, las de objeto y las de sujeto, que, según la gramática generativa, requieren diferentes operaciones de movimiento para ser interpretadas. Se han realizado al respecto decenas de experimentos con técnicas de neuroimagen; sobre el inglés, tanto con fMRI (Just et al. 1996; Cooke et al. 2002; Santi y Grodzinsky 2010), como con PET (Caplan et al. 2000 o Waters et al. 2003); sobre el hebreo con fMRI (Ben-Shachar, Hendler et al. 2003; Ben-Shachar, Palti et al. 2004); sobre el alemán, también con fMRI (Fiebach et al. 2005), entre otros muchos. La conclusión principal de todos ellos es que las subordinadas de objeto requieren más activación neuronal que las de sujeto. Sin embargo, estas técnicas son relativamente invasivas e incómodas (véase apartado 2), por lo que, en 2015, Hassanpour y sus colegas intentaron superar algunas de las carencias de la fMRI y la PET utilizando el HD-DOT. Presentaron a 10 jóvenes anglohablantes nativos unas frases con subordinadas de relativo de objeto y de sujeto cuyos agentes alternaban entre femeninos y masculinos (*los hombres que ayudan a las mujeres son útiles; las mujeres que ayudan a los hombres son útiles; los hombres a los que ayudan las mujeres son útiles; las mujeres a las que ayudan los hombres son útiles*). La tarea consistía en pulsar el botón izquierdo si la persona que realizaba la acción era mujer y el derecho si era hombre. Las tasas de acierto y el tiempo de respuesta fueron similares para los dos tipos de subordinadas. Pero, como mostraban también los experimentos con fMRI y PET, la actividad cortical no era la misma:

Durante la comprensión correcta de las subordinadas relativas de objeto, encontramos regiones cerebrales con una actividad significativamente más fuerte que se superponía a la red central de procesamiento del habla, así como regiones complementarias que no se observaban en respuesta a las oraciones más simples,

las relativas de sujeto. En general, la HD-DOT reveló una ampliación de la red de procesamiento del habla cuando aumentó la carga de procesamiento (Hassanpout et al. 2015: 323).

Por lo tanto, la respuesta a nuestra primera pregunta en este apartado vuelve a ser negativa: las oraciones subordinadas de relativo de sujeto y las de objeto se procesan de modo diferente.

Otro de los temas reiteradamente estudiados en la aplicación de técnicas de neuroimagen a cuestiones sintácticas y que nos servirá de enlace con el epígrafe 3.4, dedicado a la semántica, es si existe alguna diferencia en la activación neuronal ante verbos intransitivos (inergativos o inacusativos) y verbos transitivos o ditransitivos. En los ejemplos que encabezan este apartado encontramos un verbo intransitivo puro o inergativo, cuyo argumento tiene el papel de agente (*Ana apareció*) frente a un transitivo alternante (*Ana abrió la puerta*, que admite *la puerta se abrió*) y un transitivo no alternante (*Ana cortó el pan*, **el pan se cortó*). Es el tipo de estímulos utilizados en trabajos sobre el inglés como los de Thompson et al. (2007), den Ouden et al. (2009), Meltzer-Asscher, Schuchard et al. (2013) o Meltzer-Asscher, Mack et al. (2015). En este último se presentaron por escrito a 16 jóvenes y 13 personas mayores (54-70 años) setenta y dos verbos del inglés en infinitivo (18 inergativos o intransitivos puros, como *salivar*, que solo admiten un argumento; 18 inacusativos que solo se usan como intransitivos, como *aparecer*, cuya estructura argumental consta también de un solo elemento; 18 inacusativos que pueden usarse como transitivos, con dos argumentos, o como intransitivos, con uno: es el caso de *abrir*; y 18 transitivos, que presentan dos argumentos: *aceptar*) más 48 pseudoverbos. La tarea consistía en distinguir los verbos de los pseudoverbos mientras se tomaban imágenes de fMRI. Los resultados conductuales solo mostraron mayores tiempos de reacción en los verbos inacusativos no alternantes (*aparecer*). También la fMRI mostró que el incremento en el número de papeles temáticos provocaba activaciones en la zona posterior perisilviana del hemisferio izquierdo y que el procesamiento de estos verbos activaba un área diferente de los otros (el giro inferior frontal izquierdo), lo cual se interpreta como un reflejo de su dificultad cognitiva. En resumen, tal vez en contra de lo que hubiéramos podido suponer, este experimento nos permitiría responder que la estructura más difícil de las enunciadas más arriba es *Ana apareció*.

3.4. En el nivel semántico: ¿Se procesan igual las relaciones de semejanza que las de contigüidad? ¿Pueden los experimentos con neuroimagen aportar algo a lo que sabemos sobre la relación entre significado y referente?

En trabajos recientes sobre las bases neuronales del procesamiento semántico se afirma que la red neuronal que permite conocer los significados de las palabras es ya tan conocida, que lo único que sigue sujeto a debate es el modelo que mejor explica la función de cada una de sus partes:

Aunque la red de regiones cerebrales que sustentan la cognición semántica está bien establecida (Binder et al. 2009; Lambon Ralph et al. 2017), no está claro si las distinciones funcionales entre estos sitios reflejan diferencias en contenidos [*content*] o en procesos [*process*] (Mirman et al. (2017) en Teige et al. 2019: 308).

En el mismo sentido se pronuncia Varo Varo, que además reclama una mayor atención por parte de la Lingüística hacia esta fuente de información:

La Semántica léxica no puede desarrollarse a espaldas de los avances en neurociencia, que nos proporcionan evidencias empíricas abundantes sobre dónde, cuándo y cómo se procesa el significado de las palabras (Varo Varo 2017: 192).

Esa abundancia en las investigaciones ha permitido la publicación de bastantes meta-análisis que reutilizan los datos previos de experimentos con neuroimagen y obtienen conclusiones más generalizables: Joseph (2001), Bookheimer (2002), Thompson-Schill (2003), Damasio et al. (2004), Gerlach (2007), Binder et al. (2009), etc. En buena parte, los esfuerzos de los investigadores se han centrado en localizar de forma precisa las áreas del cerebro que se activan durante el procesamiento semántico. Para un lingüista, saber si el giro temporal anterior tiene más o menos activación que el lóbulo parietal inferior tiene solo un interés relativo. Pero seguramente sí le interesa saber es que, a diferencia de lo que ocurre con la sintaxis, o la fonología, recuperar el significado léxico requiere la participación de regiones diversas en la corteza cerebral, y no solo en las áreas atribuidas al procesamiento lingüístico; también otras, multimodales y asociativas, generalmente localizadas preferentemente en el hemisferio izquierdo, pero en un tercio de los experimentos, también en el derecho. La cita siguiente y su representación gráfica (Figura 4), procedentes de la revisión de Binder et al. (2009), pueden resumir la información que necesitamos.

Las regiones corticales implicadas en el procesamiento semántico pueden agruparse en tres grandes categorías: la corteza de asociación posterior multimodal y heteromodal, la corteza prefrontal heteromodal y las regiones límbicas mediales. La expansión de estas regiones en el cerebro humano, en comparación con el de los primates no humanos, explicaría capacidades exclusivamente humanas para utilizar el lenguaje de forma productiva, para planificar, para resolver problemas y crear artefactos culturales y tecnológicos, todo lo cual depende de una recuperación y manipulación fluidas y eficientes del conocimiento semántico (Binder et al. 2009: 2783-2784).

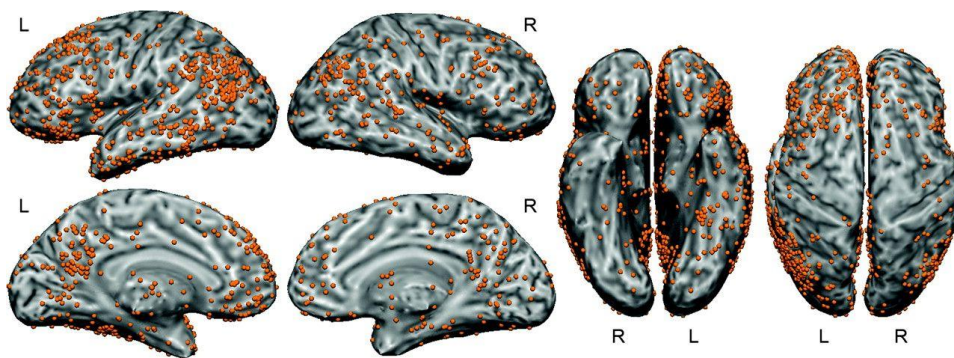


Figura 4. Binder et al. (2009: 2771). Proyección en una superficie cortical de los 1150 focos de activación publicados en los estudios revisados

Puede ser también relevante para la Lingüística saber qué rasgos semánticos tienen repercusiones neuronales, es decir, qué diferencias de significado provocan patrones de activación particulares. Y a este respecto, hay un contraste que ha sido estudiado muy concienzudamente en trabajos con neuroimagen: la diferencia entre relaciones de semejanza, basadas en características físicas compartidas (*perro-lobo*, *gato-topo*), y las

de contigüidad, basadas en coapariciones frecuentes (*perro-correa*, *topo-tierra*). Es el tipo de contraste que da pie a la tradicional diferencia entre metáforas y metonimias. Los resultados de múltiples experimentos han permitido la formulación de dos teorías para su explicación: la del *centro dual* (*dual hub*) postula que hay una zona especializada en procesar las relaciones de semejanza (los lóbulos temporales anteriores, ATL) y otra que se dedica a las de contigüidad (regiones temporoparietales, concretamente la circunvolución temporal media posterior, pMTG). En cambio, desde la teoría de *centro y radios* (*hub and spoke*) o la *Cognición Semántica Controlada* (CSC) se propone que ambas relaciones de significado se procesan en los ATL (el *centro*), pero que el papel del pMTG izquierdo, uno de los *radios*, sería recuperar esa información y, así, esta región estaría encargada del “control” semántico (una revisión en Lambon Ralph et al. 2017). En un trabajo con estimulación magnética transcraneal se puso de manifiesto la importancia de la frecuencia de uso de las palabras a este respecto: cuanto más infrecuente es la relación entre esas parejas (por ejemplo, *champagne-piscina*, frente a *champagne-fiesta*; o *dedo-guante*, frente a *dedo-garra*) mayor será la implicación de la zona radial, y más fuerte su relación con el centro, su ATL (Teige et al. 2019).

Otra fructífera línea de intersección entre la teoría semántica y la Neurolingüística, con las técnicas de neuroimagen como metodología, son las propuestas de la semántica corporeizada, ligada desde sus inicios a los circuitos cerebrales. Puede verse un rápido bosquejo en Varo Varo (2017: 1041-1042) desde las primeras formulaciones hasta otras más recientes, como la de Pulvermüller (2013)¹⁰. En la Universitat Jaume I de Castellón se ha llevado a cabo una línea de investigación en este marco, con preguntas como la siguiente: ¿qué ocurre cuando oímos o leemos palabras como *ajo*, *canela* o *vómito*? La respuesta es que se activan las áreas del cerebro necesarias para procesar los olores (González et al. 2006), como se activan las áreas gustativas cuando leemos *aceituna*, *sal* o *chorizo* (Barros-Loscertales et al. 2012). Del mismo modo, las palabras *besar*, *mascar* o *soplar* activan áreas motoras relacionadas con los músculos faciales, y *caminar* o *correr* las que dominan los músculos de las piernas (Hauk et al. 2004; Pulvermüller et al. 2005; Pulvermüller y Fadiga 2010). En un artículo divulgativo sobre los resultados de este grupo, González reflexiona sobre las repercusiones que estos descubrimientos tienen para la teoría del significado léxico:

Siempre hemos pensado en los significados como algo abstracto, amodal, desligado del terreno (*ground*) de las sensaciones y las acciones motoras. Pero lo que el cerebro nos indica es que incorpora y entreteje esta información de bajo nivel en la construcción de los elevados significados. Una información que en última instancia es corporal, tiene que ver con nuestras sensaciones y movimientos. Cuando aprendemos las palabras, además de los circuitos responsables de su forma fonológica y articulatoria, activamos en estrecha vecindad temporal otras áreas encargadas de la información sensorial y motora procedente de nuestras experiencias con el objeto referenciado (González Álvarez 2012: 129).

Los estudios sobre procesamiento lingüístico alcanzan también, aunque con menor frecuencia, los niveles textuales, discursivos y pragmáticos. No podemos extendernos en ello, pero remitimos a trabajos sobre el español, como los de Martín-Loeches et al. (2008) con fMRI, o el meta-análisis de Frestl et al. (2008).

4. Conclusiones

Las técnicas de neuroimagen están permitiendo un conocimiento cada vez más preciso de la actividad cerebral necesaria para utilizar el lenguaje, no solo en la corteza, sino también en el cerebelo, el tálamo o los ganglios basales, sin olvidar los circuitos que los conectan. Hemos pasado de una visión dicotómica, propia de la época en que las únicas evidencias procedían de cerebros lesionados, y limitada al lóbulo temporal del hemisferio izquierdo (cuya región anterior, el área motora o de Broca, estaría encargada de la producción del habla, y la posterior, el área auditiva o de Wernicke, de su descodificación) a propuestas más flexibles y complejas:

Los centros encargados del procesamiento lingüístico no son homogéneos, sino que se organizarían como una red difusa de subcentros, cada uno de los cuales sería responsable del procesamiento de componentes específicos del lenguaje. Consecuentemente, estos subcentros neuronales podrían describirse con mayor propiedad en términos de su especialización lingüística, de manera que aquellos componentes del lenguaje (fonología, semántica, sintaxis) contarían, en principio, con un correlato estructural y funcional en lo que atañe a la organización cerebral. Y lo que es aún más significativo: cada uno de estos procesos (lingüísticos) puede a su vez fraccionarse, en el sentido de que circuitos neuronales diferentes se encargarían de aspectos distintos de los mismos (Benítez-Burraco 2006: 462).

Desde sus primeras aplicaciones al estudio del procesamiento lingüístico, en los años 90, las técnicas de neuroimagen demostraron una precisión mucho mayor que la que permitían los datos de lesionados cerebrales, en los cuales además era imposible controlar los efectos de la plasticidad y reorganización cerebral. Ciertamente, tienen limitaciones: algunas derivan de las propias técnicas, por ejemplo, el acoplamiento neurovascular requiere medidas de normalización y un software especializado, en continua revisión crítica (Phillips et al. 2016); y este a su vez incorpora modelos matemáticos cuyas diferencias quizá no siempre se calibren de forma adecuada (Huneau et al. 2015). Otras están relacionadas con cuestiones metodológicas más generales, como el número de sujetos, las pruebas estadísticas o las validaciones independientes (véanse las recomendaciones de Poldrack et al. 2017).

Pese a ello, han dado lugar a “muchos hallazgos [que] han sido replicados una y otra vez hasta llevar a algunas conclusiones consistentes e indiscutibles” (Price 2012: 1053). Entre esas evidencias ya plenamente aceptadas, Price (en una revisión y síntesis personal, atípica pero enormemente interesante, de los experimentos con PET y fMRI publicados entre 1992 y 2011 para investigar la percepción y producción del lenguaje oral y la lectura) destaca la diferencia entre procesos que se localizan en estructuras específicas (por ejemplo, el procesamiento sensorial y motor) y los que surgen de un patrón de activación distribuido en áreas diferentes que participan en múltiples funciones. Por ejemplo, el procesamiento fonológico se localiza en regiones concretas, en la integración entre las áreas especializadas en procesamiento acústico (en regiones del lóbulo temporal) y las dedicadas al procesamiento articulatorio (en regiones premotoras y frontoparietales). En cambio, la semántica requiere la activación de muchas áreas en el hemisferio izquierdo, cuyos papeles aún no están suficientemente definidos.

Sin embargo, es importante destacar que la mayoría de esos esfuerzos se encaminan a lograr fines de interés clínico o neuropsicológico, no lingüístico. La localización precisa de áreas anatómicas en las que tiene lugar cada uno de los procesos y subprocesos

necesarios para usar el lenguaje es muy importante si el objetivo es el éxito en una operación de un tumor cerebral, pero determinar si para la interpretación de los significados es más importante el giro frontal inferior izquierdo o el giro angular izquierdo no aporta gran cosa a la investigación lingüística. A menudo, las respuestas a las preguntas de la Lingüística deben entresacarse de esas otras conclusiones cuyas repercusiones nos son ajenas.

Otra consecuencia de esa presencia minoritaria de lingüistas en los equipos de investigación con técnicas de neuroimagen es la reiteración en el tipo de estímulos que se utilizan y en los fenómenos que se abordan, incluso en los errores de concepto subyacentes que se heredan de unos a otros. En su revisión de las publicaciones sobre evaluación intraoperatoria (realizada mientras se interviene quirúrgicamente el cerebro), Zanin et al. (2017) concluyen:

La imagen global resultante de nuestro análisis indica que lo que se ha investigado en la bibliografía intraoperatoria es el procesamiento morfosintáctico, más que la sintaxis pura. El estudio de la neurobiología de la sintaxis durante la cirugía [con la persona] despierta, por tanto, parece estar todavía en una fase inicial, y necesita *investigaciones sistemáticas con base lingüística* (Zanin et al. 2017: 1579; el destacado en cursiva es nuestro).

El primer paso, si queremos satisfacer esa necesidad de fundamentar lingüísticamente las investigaciones sobre procesamiento del habla con técnicas de neuroimagen, es conseguir que la formación de los lingüistas incluya las bases necesarias para comprender y analizar los resultados de este tipo de trabajos, de modo que contemos con suficiente masa crítica. Además, es necesaria la integración de lingüistas en los equipos multidisciplinares que ponen en práctica esos estudios, con el fin de controlar las variables lingüísticas en la metodología y asegurar que los diseños experimentales están libres de errores como los mencionados anteriormente. La meta última sería lograr que la planificación completa de los estudios, y sus objetivos, respondan a intereses provenientes de la teoría lingüística. Es alentador comprobar que eso ya está sucediendo; un ejemplo es el trabajo de Tanaka et al. (2019) que demuestra la trascendencia de la operación *merge*, propuesta por el Programa Minimista, utilizando la fMRI: sus autores pertenecen a facultades de Lingüística. Contamos con investigadores destacados, tanto en el ámbito internacional como en el nacional, que ya han abierto el camino; esperemos que su número se incremente exponencialmente en los próximos años, lo cual conllevará, sin duda, un gran enriquecimiento mutuo de la Lingüística y la ciencia cognitiva.

5. Agradecimientos

Agradezco al prof. Michael Winterdahl, coordinador de la *Interdisciplinary Summer School on Neuroimaging – 2019 with special focus on imaging in research*, de Aarhus University, y a todo el equipo docente de la misma, la oportunidad de realizar el curso y la formación recibida para preparar este trabajo; por supuesto, también a los organizadores de Psylex V el haberme colocado ante al reto de llevarlo a cabo, y a los revisores y editoras sus sugerencias, que sin duda han mejorado el texto inicial. Cualquier error en el mismo es responsabilidad exclusiva de la autora.

6. Referencias

- Adank, Patti; Nuttall, Helen E.; Banks, Briony; Kennedy-Higgins, Daniel. 2015. Neural bases of accented speech perception. *Frontiers in human neuroscience* 9: 558.
- Armony, Jorge L.; Trejo-Martínez, David; Hernández, Dailett. 2012. Resonancia magnética funcional (RMf): principios y aplicaciones en neuropsicología y neurociencias cognitivas. *Neuropsicología Latinoamericana* 4.2: 36-50.
- Aziz-Zadeh, Lisa; Damasio, Antonio. 2008. Embodied semantics for actions: Findings from functional brain imaging. *Journal of Physiology-Paris* 102.1-3: 35-39.
- Bambini, Victoria; Canal, Paolo. 2021. Neurolinguistic Research on the Romance Languages. *Oxford Research Encyclopedia of Linguistics*. DOI: 10.1093/acrefore/9780199384655.013.443
- Barros-Loscertales, Alfonso; González, Julio; Pulvermüller, Friedemann; Ventura-Campos, Noelia, Bustamante, Juan Carlos; Costumero, Víctor; Parcet, María Antonia; Ávila, César. 2012. Reading salt activates gustatory brain regions: fMRI evidence for semantic grounding in a novel sensory modality. *Cerebral Cortex*, 22.11: 2554-2563.
- Barsalou, Lawrence W. 2010. Grounded cognition: Past, present, and future. *Topics in cognitive science* 2.4: 716-724.
- Belyk, Michel; Brown, Steven. 2014. Perception of affective and linguistic prosody: an ALE meta-analysis of neuroimaging studies. *Social cognitive and affective neuroscience* 9.9: 1395-1403.
- Ben-Shachar, Michal; Hendler, Talma; Kahn, Itamar; Ben Bashat, Dafna; Grodzinsky, Yosef. 2003. The neural reality of syntactic transformations: Evidence from functional magnetic resonance imaging. *Psychological science* 14.5: 433-440.
- Ben-Shachar, Michal; Palti, Dafna; Grodzinsky Yosef. 2004. Neural correlates of syntactic movement: Converging evidence from two fMRI experiments. *NeuroImage* 21: 1320-1336.
- Benítez-Burraco, Antonio. 2006. Caracterización neuroanatómica y neurofisiológica del lenguaje humano. *Revista Española de Lingüística* 35.2: 461-494.
- Benítez-Burraco, Antonio; Murphy, Elliot. 2019. Why brain oscillations are improving our understanding of language. *Frontiers in behavioral neuroscience* 13: 190.
- Beres, Anna M. 2017. Time is of the essence: A review of electroencephalography (EEG) and event-related brain potentials (ERPs) in language research. *Applied psychophysiology and biofeedback* 42.4: 247-255.
- Bernal, Byron; Ardila, Alfredo. 2016. From hearing sounds to recognizing phonemes: Primary auditory cortex is a truly perceptual language area. *AIMS Neur* 3.4: 454-473.
- Binder, Jeffrey R.; Desai, Rutvik H.; Graves, William W.; Conant, Lisa L. 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral cortex* 19.12: 2767-2796.
- Bookheimer, Susan. 2002. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annual review of neuroscience* 25.1: 151-188.
- Brunellière, Angèle; Dufour, Sophie; Nguyen, Noël. 2011. Regional differences in the listener's phonemic inventory affect semantic processing: a mismatch negativity (MMN) study. *Brain and Language* 117.1: 45-51.
- Caplan, David. 2001. Functional neuroimaging studies of syntactic processing. *Journal of psycholinguistic research* 30.3: 297-320.

- Caplan, David; Alpert, Nathaniel; Waters, Gloria; Olivieri, Anthony. 2000. Activation of Broca's area by syntactic processing under conditions of concurrent articulation. *Human brain mapping* 9.2: 65-71.
- Carreiras, Manuel; Carr, Lindsay; Barber, Horacio A.; Hernandez, Arturo E. 2010. Where syntax meets math: Right intraparietal sulcus activation in response to grammatical number agreement violations. *Neuroimage* 49.2: 1741-1749.
- Chiao, Joan Y. (ed.). 2009. *Cultural neuroscience: Cultural influences on brain function*. Elsevier.
- Chiao, Joan. Y.; Cheon, Bobby K; Pornpattanakul, Narun; Mrazek, Alissa J.; Blizinsky, Katherine D. 2013. Cultural neuroscience: progress and promise. *Psychological inquiry* 24.1: 1-19.
- Chomsky, Noam. 1957. *Syntactic Structures*. Paris: Mouton.
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge: MIT Press.
- Chomsky, Noam; Halle, Morris. 1968. *The sound pattern of English*. New York: Harper & Row.
- Cooke, Ayanna; Zurif, Edgar B.; DeVita, Christian; Alsop, David; Koenig, Phyllis; Detre, John; Gee, James; Pinango, Maria; Balogh, Jennifer; Grossman, Murray. 2002. Neural basis for sentence comprehension: Grammatical and short-term memory components. *Human brain mapping* 15.2: 80-94.
- Corrales-Quispircra, Carmen; Gadea, M. Engracia; Espert, Raúl. 2020. Estimulación de corriente continua transcraneal e intervención logopédica en personas con afasia: revisión sistemática de la bibliografía. *Rev Neurol* 70: 351-64.
- Damasio, Antonio R.; Tranel, Daniel. 1993. Nouns and verbs are retrieved with differently distributed neural systems. *Proceedings of the National Academy of Sciences* 90.11: 4957-4960.
- Damasio, Hanna; Tranel, D.; Grabowski, T. J.; Adolphs, R.; Damasio, Antonio. 2004. Neural systems behind word and concept retrieval. *Cognition* 92.1-2: 179-229.
- den Ouden, Dirk-Bart; Fix, Steve; Parrish, Todd B.; Thompson, Cynthia K. 2009. Argument structure effects in action verb naming in static and dynamic conditions. *Journal of Neurolinguistics* 22.2: 196-215.
- Dick, Antony S.; Bernal, Byron; Tremblay, Pascale. 2014. The language connectome: new pathways, new concepts. *The Neuroscientist* 20.5: 453-467.
- Dikker, Suzanne; Assaneo, M Florencia; Gwilliams, Laura; Wang, Lin; Kösem, Anne. 2020. Magnetoencephalography and language. *Neuroimaging Clinics* 30.2: 229-238.
- Dronkers, Nina F; Plaisant, Odile; Iba-Zizen, M. T.; Cabanis, Emmanuel Alain. 2007. Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong. *Brain* 130.5: 1432-1441.
- Erdocia, Kepa; Laka, Itziar. 2018. Negative transfer effects on L2 word order processing. *Frontiers in psychology* 9: 337.
- Feldman, Jerome; Narayanan, Srinivas. 2004. Embodied meaning in a neural theory of language. *Brain and language* 89.2: 385-392.
- Ferstl, Evelyn C.; Neumann, Jane; Bolger, Carsten; von Cramon, D. Yves. 2008. The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Human brain mapping* 29.5: 581-593.
- Fiebach, Christian J.; Schlwewsky, Matthias; Lohmann, Gabriele; von Cramon, D. Y.; Friederici, Angela D. 2005. Revisiting the role of Broca's area in sentence processing: syntactic integration versus syntactic working memory. *Human brain mapping* 24.2: 79-91.

- Fitzgerald, Des; Callard, Felicity. 2015. Social science and neuroscience beyond interdisciplinarity: Experimental entanglements. *Theory, Culture & Society* 32.1: 3-32.
- Friederici, Angela D.; Chomsky, Noam; Berwick, Robert C.; Moro, Andrea; Bolhuis, Johan J. 2017. Language, mind and brain. *Nature Human Behaviour* 1.10: 713-722.
- Frye, Richard E.; Rezaie, Roozbeh; Papanicolaou, Andrew. C. 2009. Functional neuroimaging of language using magnetoencephalography. *Physics of life reviews* 6.1: 1-10.
- Galantucci, Bruno; Fowler, Carol A.; Turvey, Michael T. 2006. The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review* 13: 361-377.
- García-Pentón, Lorna; Pérez Fernández, Alejandro; Iturria-Medina, Yasser; Gillon-Dowens, Margaret; Carreiras, Manuel. 2014. Anatomical connectivity changes in the bilingual brain. *Neuroimage* 84: 495-504.
- Gerlach, Christian. 2007. A review of functional imaging studies on category specificity. *Journal of Cognitive Neuroscience* 19.2: 296-314.
- Giménez-Roldán, Santiago. 2017. Una revisión crítica sobre la contribución de Broca a la afasia: desde la prioridad al sombrerero Leborgne. *Neurosciences and History* 5.2: 58-68.
- González Álvarez, Julio. 2012. La representación neurobiológica de las palabras y sus significados en el cerebro. *Salud y Ciencia* 19: 126-9. <https://www.siicsalud.com/des/expertoimpreso.php/126104>
- González, Julio; Barros-Loscertales, Alfonso; Pulvermüller, Friedemann; Meseguer, Vanessa; Sanjuán, Ana; Belloch, Vicente; Ávila, César. 2006. Reading cinnamon activates olfactory brain regions. *Neuroimage* 32.2: 906-912.
- González, Rafael; Hornauer-Hughes, Andrea. 2014. Cerebro y lenguaje. *Revista Hospital Clínico Universidad de Chile* 25.1: 144-153.
- Goslin, Jeremy; Duffy, Hester; Floccia, Caroline. 2012. An ERP investigation of regional and foreign accent processing. *Brain and language* 122.2: 92-102.
- Grodzinsky, Yosef; Friederici, Angela D. 2006. Neuroimaging of syntax and syntactic processing. *Current opinion in neurobiology* 16.2: 240-246.
- Hämäläinen, Sini; Sairanen, Viljami; Leminen, Alina; Lehtonen, Minna. 2017. Bilingualism modulates the white matter structure of language-related pathways. *NeuroImage* 152: 249-257.
- Hari, Riita; Puce, Aina. 2017. *MEG-EEG Primer*. Oxford: Oxford University Press.
- Haro, Juan. 2022. Una introducción al uso de los potenciales evocados en el estudio del lenguaje. *Estudios de Lingüística del Español* 45: 185-204.
- Hassanpour, Mahlega S.; Eggebrecht, Adam T.; Culver, Joseph P.; Peelle, Jonathan E. 2015. Mapping cortical responses to speech using high-density diffuse optical tomography. *NeuroImage* 117: 319-326.
- Hauk, Olaf. 2016. What does it mean? A review of the neuroscientific evidence for embodied lexical semantics. En G. Hicko y S. L. Small, eds. *Neurobiology of language*. Academic press, pp. 777-788.
- Hauk, Olaf; Johnsrude, Ingrid; Pulvermüller, Friedemann. 2004. Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 22: 301-307.
- Hauk, Olaf; Tschentscher, Nadja. 2013. The body of evidence: what can neuroscience tell us about embodied semantics? *Frontiers in psychology* 4: 50.
- Hernandez, Arturo E.; Kotz, Sonja A.; Hofmann, Juliane; Valentin, Vivian V.; Dapretto, Mirella; Bookheimer, Susan Y. 2004. The neural correlates of grammatical gender decisions in Spanish. *Neuroreport* 15.5: 863-866.

- Hernández, Mireia; Ventura-Campos, Noelia; Costa, Albert; Miró-Padilla, Anna; Ávila, César. 2019. Brain networks involved in accented speech processing. *Brain and language* 194: 12-22.
- Holmes, Emma; Zeidman, Peter; Friston, Karl J; Griffiths, Timothy D. 2021. Difficulties with speech-in-noise perception related to fundamental grouping processes in auditory cortex. *Cerebral Cortex* 31.3: 1582-1596.
- Horwitz, Barry; Poeppel, David. 2002. How can EEG/MEG and fMRI/PET data be combined? *Human brain mapping* 17.1: 1-3.
- Horwitz, Barry; Wise, Richard J. 2008. PET research of language. En Brigitte Stemmer y Harry A. Whitaker, eds. *Handbook of the neuroscience of language*. Academic Press, pp. 71-80.
- Huneau, Clément; Benali, Habib; Chabriat, Hugues. 2015. Investigating human neurovascular coupling using functional neuroimaging: a critical review of dynamic models. *Frontiers in neuroscience* 9: 467.
- Immordino-Yang, Mary Helen. 2013. Studying the effects of culture by integrating neuroscientific with ethnographic approaches. *Psychological Inquiry* 24.1: 42-46.
- Jakobson, Roman; Fant, Gunnar; Halle, Morris. 1952. *Preliminaries to speech analysis. The distinctive features and their correlates*. Cambridge: The MIT Press
- Johnsrude, Ingrid S.; Giraud, Anne Lise; Frackowiak, Richard S. 2002. Functional imaging of the auditory system: the use of positron emission tomography. *Audiology and Neurotology* 7.5: 251-276.
- Joseph, Jane E. 2001. Functional neuroimaging studies of category specificity in object recognition: a critical review and meta-analysis. *Cogn Affect Behav Neurosci* 1: 119-136.
- Just, Marcel Adam; Carpenter, Patricia A.; Keller, Timothy A.; Eddy, William; Thulborn, Keith. 1996. Brain activation modulated by sentence comprehension. *Science* 274.5284: 114-116.
- Kanero, Junko; Imai, Mutsumi; Okuda, Jiro; Okada, Hiroyuki; Matsuda, Tetsuya. 2014. How sound symbolism is processed in the brain: a study on Japanese mimetic words. *PLoS one* 9.5: e97905.
- Kekic, Maria; Boysen, Elena; Campbell, Iain C.; Schmidt, Ulrike. 2016. A systematic review of the clinical efficacy of transcranial direct current stimulation (tDCS) in psychiatric disorders. *Journal of psychiatric research* 74: 70-86.
- Kemmerer, David. 2014. Word classes in the brain: Implications of linguistic typology for cognitive neuroscience. *Cortex* 58: 27-51.
- La Vaque, T. J. 1999. The history of EEG Hans Berger. *Journal of Neurotherapy: Investigations in Neuromodulation, Neurofeedback and Applied Neuroscience* 3.2: 1-9.
- Lambon Ralph, Matthew; Jefferies, Elizabeth; Patterson, Karalyn; Rogers, Timothy. 2017. The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience* 18.1: 42-55.
- Lieberman, Alvin M.; Mattingly, Ignatius G. 1985. The motor theory of speech perception revised. *Cognition* 21.1: 1-36.
- Lizarazu, Mikel. 2017. *Speech-brain synchronization: a possible cause for developmental dyslexia*. Tesis Doctoral. Universidad del País Vasco-Euskal Herriko Unibertsitatea.
- Ljungberg, E.; Damestani, Nikou L.; Wood, Tobias C.; Lythgoe, David J.; Zelaya, Fernando; Williams, Steven C.R.; Solana, Ana Beatriz; Barker, Gareth J.; Wiesinger, Florian. 2021. Silent zero TE MR neuroimaging: Current state-of-the-

- art and future directions. *Progress in Nuclear Magnetic Resonance Spectroscopy* 123: 73-93.
- Martín-Loeches, Manuel; Casado, Pilar; Hernández-Tamames, Juan A.; Álvarez-Linera, Juan. 2008. Brain activation in discourse comprehension: a 3t fMRI study. *Neuroimage* 41.2: 614-622.
- Meltzer-Asscher, Aya; Mack, Jennifer E.; Barbieri, Elena; Thompson, Cynthia K. 2015. How the brain processes different dimensions of argument structure complexity: Evidence from fMRI. *Brain and language* 142: 65-75.
- Meltzer-Asscher, Aya; Schuchard, Julia; den Ouden, Dirk-Bart; Thompson, Cynthia K. 2013. The neural substrates of complex argument structure representations: Processing “alternating transitivity” verbs. *Language and cognitive processes* 28.8: 1154-1168.
- Mercadillo, Roberto E.; Díaz, José Luis. 2013. Neuroscience and ethnography: an interdisciplinary revision and a cognitive proposal based on compassion research in Mexico. *International journal of psychological research* 6: 94-108.
- Miniussi, Carlo; Cotelli, Maria; Manenti, Rosa. 2010. Transcranial Magnetic Stimulation in the Study of Language and Communication. En M. Balconi, ed. *Neuropsychology of Communication*. Milán: Springer, pp. 47-59.
- Mitsuhashi, Takumi; Sugano, Hidenori; Asano, Keiko; Nakajima, Takayuki; Nakajima, Madoka; Okura, Hidehiro; Iimura, Yasushi, Suzuki, Hiroharu; Tange, Yuichi; Tanaka, Toshihisa; Aoki, Shigeki; Arai, Hajime. 2020. Functional MRI and structural connectome analysis of language networks in Japanese-English bilinguals. *Neuroscience* 431: 17-24.
- Mohades, Seyede Ghazal; Struys, Esli; Van Schuerbeek, Peter; Mondt, Katrien; Van De Craen, Piet; Luypaert, Robert. 2012. DTI reveals structural differences in white matter tracts between bilingual and monolingual children. *Brain Research* 1435: 72-80.
- Molinaro, Nicola; Barber, Horacio A.; Pérez, Alejandro; Parkkonen, Lauri; Carreiras, Manuel. 2013. Left fronto-temporal dynamics during agreement processing: Evidence for feature-specific computations. *NeuroImage* 78: 339-352.
- Mompeán, José A.; Fregier, Amandine; Valenzuela, Javier. 2020. Iconicity and systematicity in phonaestemes: A cross-linguistic study. *Cognitive Linguistics* 31.3: 515-548.
- Monti, Alessia; Ferrucci, Roberta; Fumagalli, Manuela; Mameli, Francesca; Cogiamanian, Filippo; Ardolino, Gianluca; Prioriet, Alberto. 2013. Transcranial direct current stimulation (tDCS) and language. *Journal of Neurology, Neurosurgery & Psychiatry* 84.8: 832-842.
- Murphy, Elliot. 2018. Interfaces (travelling oscillations) + recursion (delta-theta code) = language. En E. M. Luef y M. Marín, eds. *The Talking Species: Perspectives on the Evolutionary, Neuronal and Cultural Foundations of Language*. Graz: Unipress Graz Verlag: 251-269
- Murphy, Elliot. 2020. *The oscillatory nature of language*. Cambridge: Cambridge University Press.
- Nuttall, Helen E.; Kennedy-Higgins, Dan; Devlin, Joseph T.; Adank, Patti. 2018. Modulation of intra-and inter-hemispheric connectivity between primary and premotor cortex during speech perception. *Brain and language* 187: 74-82.
- Ohta, Shinri; Fukui, Naoki; Sakai, Kuniyoshi L. 2013. Computational principles of syntax in the regions specialized for language: integrating theoretical linguistics and functional neuroimaging. *Frontiers in behavioral neuroscience* 7: 204.

- Palm, Ulrich; Hasan, Alkomiet; Strube, Wolfgang; Padberg, Frank. 2016. tDCS for the treatment of depression: a comprehensive review. *European archives of psychiatry and clinical neuroscience* 266.8: 681-694.
- Paredes Duarte, María Jesús; Varo Varo, Carmen. 2006. Lenguaje y cerebro: conexiones entre neurolingüística y psicolingüística. En B. Gallardo, C. Hernández y V. Moreno, eds. *Lingüística clínica y neuropsicología cognitiva. Actas del Primer Congreso Nacional de Lingüística Clínica. Vol 1: Investigación e intervención en patologías del lenguaje*. Valencia: Universitat, pp. 107-119.
- Peelle, Jonathan E. 2017. Optical neuroimaging of spoken language. *Language, cognition and neuroscience* 32.7: 847-854.
- Phillips, Aaron A.; Chan, Franco Hn.; Zi Zheng, Mei Mu; Krassioukov, Andrei V.; Ainslie, Philip. 2016. Neurovascular coupling in humans: physiology, methodological advances and clinical implications. *Journal of Cerebral Blood Flow & Metabolism* 36.4: 647-664.
- Pliatsikas, Christos; Meteyard, Lotte; Veríssimo, João; DeLuca, Vincent; Shattuck, Kyle; Ullman, Michael T. 2020. The effect of bilingualism on brain development from early childhood to young adulthood. *Brain Structure and Function* 225.7: 2131-2152.
- Poldrack, Russell A.; Baker, Chris I.; Durnez, Joke; Gorgolewski, Krzysztof J.; Matthews, Paul M.; Munafò, Marcus R.; Nichols, Thomas E.; Poline, Jean-Baptiste; Vul, Edward; Yarkoni, Tal. 2017. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature reviews neuroscience* 18.2: 115-126.
- Przeździk, Izabela; Haak, Koen V.; Beckmann, Christian F.; Bartsch, Andreas. 2019. The human language connectome. En P. Hagoort, ed. *Human language: From genes and brains to behavior*. Cambridge, MA: The MIT Press, pp. 467-480.
- Price, Cathy J. 2010. The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences* 1191.1: 62-88.
- Price, Cathy J. 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62.2: 816-847.
- Pulvermüller, Friedemann. 2013. How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in cognitive sciences* 17.9: 458-470.
- Pulvermüller, Friedemann; Fadiga Luciano. 2010. Active perception: sensorimotor circuits as a cortical basis for language. *Nat Rev Neurosci* 11: 351-360.
- Pulvermüller, Friedemann; Shtyrov, Yuri; Ilmoniemi, Risto. 2005. Brain signatures of meaning access in action word recognition. *J Cogn Neurosci* 17: 884-892.
- Quaresima, Valentina; Bisconti, Silvia; Ferrari, Marco. 2012. A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. *Brain and language* 121.2: 79-89.
- Raichle, Marcus E. 2009. A brief history of human brain mapping. *Trends in neurosciences* 32.2: 118-126.
- Rampinini, Alessandra C.; Ricciardi, Emiliano. 2017. In favor of the phonemic principle: a review of neurophysiological and neuroimaging explorations into the neural correlates of phonological competence. *Studi e Saggi Linguistici* 55.1: 95-123.
- Ramsey, Nick F.; Salari, Efraim; Aarnoutse, Erik J.; Vansteensel, Mariska; Bleichner, Martin Georg; Freudenburg Z. V. 2018. Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. *Neuroimage* 180: 301-311.

- Revill, Kate Pirog; Namy, Laura L.; Clepper DeFife, Lauren; Nygaard, Lynne C. 2014. Cross-linguistic sound symbolism and crossmodal correspondence: Evidence from fMRI and DTI. *Brain and Language* 128.1: 18-24.
- Rizzolatti, Giacomo; Fadiga, Luciano; Gallese, Vittorio; Fogassi, Leonardo. 1996. Premotor cortex and the recognition of motor actions. *Cognitive brain research* 3.2: 131-141.
- Rogalsky, Corianne; Love, Tracy; Driscoll, David; Anderson, Steven W.; Hickok, Gregory. 2011. Are mirror neurons the basis of speech perception? Evidence from five cases with damage to the purported human mirror system. *Neurocase* 17.2: 178-187.
- Rogers, Jack C.; Möttönen, Riikka; Boyles, Rowan; Watkins, Kate E. 2014. Discrimination of speech and nonspeech sounds following theta-burst stimulation of the motor cortex. *Frontiers in Psychology* 5: 754.
- Santi, Andrea; Grodzinsky, Yosef. 2010. fMRI adaptation dissociates syntactic complexity dimensions. *Neuroimage* 51.4: 1285-1293.
- Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Paris: Payot.
- Schomers, Malte R.; Pulvermüller, Friedemann. 2016. Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. *Frontiers in human neuroscience* 10: 435.
- Sekiyama, Kaoru; Kanno, Iwao; Miura, Shuichi; Sugita, Yoichi. 2003. Auditory-visual speech perception examined by fMRI and PET. *Neuroscience research* 47.3: 277-287.
- Seligman, Rebecca; Brown, Ryan A. 2010. Theory and method at the intersection of anthropology and cultural neuroscience. *Social cognitive and affective neuroscience* 5.2-3: 130-137.
- Scherer, Liliane C.; Kahlaoui, Karima; Ansaldo, Ana I. 2009. Espectrografía funcional de infrarrojo cercano (fNIRS): la técnica y su aplicación en estudios de lenguaje. *Neuropsicología Latinoamericana* 1.1: 57-62.
- Stasenka, Alena; Garcea, Frank E.; Mahon, Bradford Z. 2013. What happens to the motor theory of perception when the motor system is damaged? *Language and cognition* 5.2-3: 225-238.
- Tanaka, Kyohei; Nakamura, Issa; Ohta, Shinri; Fukui, Noaki; Zushi, Mihoko; Narita, Hiroki; Sakai, Kuniyoshi L. 2019. Merge-generability as the key concept of human language: evidence from neuroscience. *Frontiers in psychology* 10: 2673.
- Teige, Catarina; Cornelissen, Piers L.; Mollo, Giovanna; Gonzalez Alam, Tirso Rene del Jesus; McCarty, Kristofor; Smallwood, Jonathan; Jefferies, Elizabeth. 2019. Dissociations in semantic cognition: Oscillatory evidence for opposing effects of semantic control and type of semantic relation in anterior and posterior temporal cortex. *Cortex* 120: 308-325.
- Thompson-Schill, Sharon L. 2003. Neuroimaging studies of semantic memory: inferring “how” from “where”. *Neuropsychologia* 41.3: 280-292.
- Thompson, Cynthia K.; Bonakdarpour, Borna; Fix, Stephen C.; Blumenfeld, Henrike K.; Parrish, Todd B.; Gitelman, Darren; Mesulam, M. Marsel. 2007. Neural correlates of verb argument structure processing. *Journal of Cognitive Neuroscience* 19.11: 1753-1767.
- Tremblay, Pascale; Dick, Anthony S. 2016. Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and language* 162: 60-71.
- Vansteensel, Mariska J.; Pels, Elmar G.M.; Bleichner, Martin G.; Branco, Mariana P.; Denison, Timothy; Freudenburg, Zachary V.; Gosselaar, Peter; Leinders, Sacha; Ottens, Thomas H.; Van Den Boom, Max A.; Van Rijen, Peter C.; Aarnoutse, Erik

- J.; Ramsey, Nick F. 2016. Fully Implanted Brain-Computer Interface in a Locked-In Patient with ALS. *N Engl J Med* 375: 2060-2066.
- Varo Varo, Carmen. 2017. Nuevos retos en la investigación del contenido léxico: elementos para una neurosemántica. *Rilce: Revista de Filología Hispánica* 33.3: 1032-1059.
- Vassal, François; Schneider, Fabien; Boutet, Claire; Jean, Betty; Sontheimer, Anna; Lemaire, Jean-Jacques. 2016. Combined DTI tractography and functional MRI study of the language connectome in healthy volunteers: extensive mapping of white matter fascicles and cortical activations. *PloS one* 11.3: e0152614.
- Vigliocco, Gabriella; Vinson, David P.; Druks, Judit; Barber, Horacio; Cappa, Stefano F. 2011. Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience & Biobehavioral Reviews* 35.3: 407-426.
- Wang, Lin; Schoot, Lotte; Brothers, Trevor; Alexander, Edward; Warnke, Lena; Minjae, Kim; Khan, Sheraz; Hamalainen, Matti; Kuperberg, Gina R. 2021, en revisión. Dynamic predictive coding across the left fronto-temporal language hierarchy: Evidence from MEG, EEG and fMRI. *bioRxiv*.
- Waters, Gloria; Caplan, David; Alpert, Nathaniel; Stanczak, Louise. 2003. Individual differences in rCBF correlates of syntactic processing in sentence comprehension: effects of working memory and speed of processing. *NeuroImage* 19.1: 101-112.
- Watkins, Kate E.; Strafella, Antonio P.; Paus, Tomás. 2003. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41.8: 989-994.
- Westwood, Samuel J.; Olson, Andrew; Miall, Chris R.; Nappo, Raffaele; Romani, Cristina. 2017. Limits to tDCS effects in language: failures to modulate word production in healthy participants with frontal or temporal tDCS. *Cortex* 86: 64-82.
- Wilson, Stephen M.; Iacoboni, Marco. 2006. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *NeuroImage* 33.1: 316-325.
- Wilson, Stephen M.; Saygin, Ayse Pinar; Sereno, Martin I.; Iacobini, Marco. 2004. Listening to speech activates motor areas involved in speech production. *Nature neuroscience* 7.7: 701-702.
- Yuan, Zhen; Ye, JongChul. 2013. Fusion of fNIRS and fMRI data: identifying when and where hemodynamic signals are changing in human brains. *Frontiers in human neuroscience* 7: 676.
- Zanin, Elia; Riva, Marco; Bambini, Valentina; Cappa, Stefano F.; Magrassi, Lorenzo; Moro, Andrea. 2017. The contribution of surgical brain mapping to the understanding of the anatomo-functional basis of syntax: A critical review. *Neurological Sciences* 38.9: 1579-1589.
- Zawiszewski, Adam; Laka, Itziar. 2020. Bilinguals processing noun morphology: Evidence for the Language Distance Hypothesis from event-related potentials. *Journal of Neurolinguistics* 55: 100908.

¹ En uno de esos increíbles casos en la historia de la ciencia, ese cerebro se ha conservado hasta nuestros días, lo cual permitió su reanálisis con modernas técnicas de resonancia magnética (Dronkers et al. 2007).

² Calificadas así porque permiten observar las regiones del cerebro que están activas (funcionando) en el momento del registro (Raichle 2009).

³ Espectroscopía funcional del infrarrojo cercano (fNIRS), y tomografía óptica difusa de alta densidad, (HD-DOT, por sus siglas en inglés); más adelante se describirán brevemente, pero puede verse un resumen de sus características en Peelle (2017).

⁴ Por ejemplo, los trabajos de Chiao (2009), Seligman y Brown (2010), Chiao et al. (2013), Mercadillo y Díaz (2013), Immordino-Yang (2013), Fitzgerald y Callard (2015), etc.

⁵ Entre los lingüistas españoles que trabajan con esta técnica está Itziar Laka, en la Universidad del País Vasco, con una línea sobre procesamiento en bilingües (Zawiszewski y Laka 2020; Erdocia y Laka 2018; etc.).

⁶ Son sustancias que se desintegran al poco de su creación: el oxígeno ^{15}O , tarda un par de minutos en hacerlo, mientras que la fluordesoxiglucosa, F^{18} , llega a permanecer casi dos horas.

⁷ Como muestra de ese predominio, los estudios publicados utilizando esta metodología en un solo año permitieron la realización de un meta-análisis sobre más de cien referencias (Price 2010).

⁸ Al principio era de medio tesla (T), pero desde 2010 hay máquinas de 7T; las más habituales son de 3T, unas 60000 veces el campo magnético de la Tierra.

⁹ Como el francés del sur frente al estándar (Brunellière et al. 2011), el inglés del sudeste de Inglaterra frente al de Gales, Yorkshire o el inglés como lengua extranjera (Goslin et al. 2012). Puede verse una revisión en Adank et al. (2015).

¹⁰ Algunas referencias interesantes a este respecto son Feldman y Narayanan (2004), Aziz-Zadeh y Damasio (2008), Hauk y Tschentscher (2013), o la revisión de Hauk (2016), entre otras.

Una introducción al uso de los potenciales evocados en el estudio del lenguaje

Juan Haro
Universitat Rovira i Virgili
juan.haro@urv.cat

Resumen

La técnica de potenciales evocados consiste en el registro de la actividad eléctrica cerebral generada tras la presentación de un estímulo. Gracias a su elevada resolución temporal, del orden de milisegundos, esta técnica es útil para analizar con precisión el curso temporal del procesamiento del lenguaje. La investigación con potenciales evocados ha logrado identificar algunos patrones de actividad cerebral asociados a distintos niveles del procesamiento y representación del lenguaje. Entre los más estudiados se encuentran el potencial de disparidad (MMN), asociado a la fonología, y los componentes N400 y P600, asociados a la semántica y sintaxis, respectivamente.

Palabras clave: potenciales evocados, lenguaje, EEG, N400, P600, potencial de disparidad.

Abstract

The event-related potential technique records the electrical brain activity generated by the presentation of a stimulus. Because of its high temporal resolution, in the order of milliseconds, this technique is useful for an accurate analysis of the time course of language processing. Research with event-related potentials has revealed patterns of brain activity associated with different levels of language processing and representation. Among the most studied, we find the mismatch negativity (MMN), associated with phonology, and the N400 and P600 components, associated with semantics and syntax, respectively.

Keywords: event-related potentials, language, EEG, N400, P600, mismatch negativity.

1. Introducción

Un problema frecuente cuando planificamos y diseñamos una investigación es escoger una medida que sea lo suficientemente fiable y válida. Existen diferentes medidas para el estudio del lenguaje. Las más empleadas y con mayor tradición en este campo son las medidas conductuales (Igoa 2022). Estas medidas se obtienen de la respuesta emitida por un participante en una tarea experimental; por ejemplo, el tiempo que tarda un participante en reconocer una palabra en una tarea de discriminación de palabras (p.ej., *pelota*) y falsas palabras (p.ej., *judapo*).

Las medidas conductuales son de gran utilidad para el estudio del lenguaje y han permitido arrojar luz sobre los mecanismos que subyacen al procesamiento y producción del lenguaje (véase Garrod 2006, para una breve revisión). Sin embargo, presentan algunas limitaciones. Una crítica frecuente es que la respuesta conductual del participante es solo el *output* de un complejo proceso. Las medidas conductuales revelan una imagen

estática de procesos dinámicos, de breve duración, compuestos por un número indeterminado de subprocesos en continua interacción. Además, la respuesta conductual puede estar influida por variables ajenas al procesamiento de estímulos lingüísticos, como son todas aquellas relacionadas con la planificación y ejecución de dicha respuesta. Un ejemplo de ello lo podemos encontrar en Balota y Abrams (1995), quienes realizaron un estudio de decisión léxica en el cual examinaron cómo afecta la manipulación de la frecuencia de una palabra a la fuerza, velocidad y duración de la respuesta motora que emiten los participantes. Los participantes disponían de una palanca con la que tenían que indicar si el estímulo presentado era una palabra o bien una falsa palabra. Debían mover la palanca hacia un lado si se trataba de una palabra o hacia el opuesto si se trataba de una falsa palabra. Los resultados mostraron que la frecuencia de la palabra no solo influyó en los tiempos de reconocimiento de las palabras, sino también en la fuerza, velocidad y duración de la respuesta motora. Los participantes movieron la palanca con mayor velocidad y fuerza al responder a palabras de alta frecuencia que al hacerlo con palabras de baja frecuencia.

Una forma de salvar las limitaciones de las medidas conductuales es complementarlas con otras medidas que permitan realizar un seguimiento preciso del desarrollo de los procesos cognitivos que preceden (y siguen) a la respuesta del participante. La técnica de potenciales evocados satisface perfectamente esta necesidad. Esta es una técnica neurofisiológica, basada en el registro de la actividad electroencefalográfica (EEG), que proporciona una medida en tiempo real de la actividad eléctrica del cerebro mediante métodos no invasivos. El presente trabajo se inicia con un breve repaso sobre el origen y el desarrollo de la técnica de potenciales evocados, seguido de una descripción de algunos componentes de potenciales evocados estrechamente relacionados con el estudio del lenguaje. Por último, se incluye un ejemplo práctico sobre cómo preparar, administrar y analizar un experimento de potenciales evocados.

2. Electroencefalografía

La EEG humana fue descubierta por el médico y psiquiatra Hans Berger (véase Figura 1). La primera publicación en la que se describe el uso de esta técnica en humanos data de 1929 (Berger 1929). En dicho trabajo, Berger explica cómo colocó un conjunto de electrodos sobre la cabeza de un participante, que a su vez estaba conectado a un galvanómetro¹, con el cual registró una curva de fluctuación constante. En el registro pudo apreciar dos tipos de ondas eléctricas: unas de mayor longitud y con una duración de 90 milisegundos (las cuales denominó *ondas de primer orden*) y otras de menor longitud y con una duración de 35 milisegundos (*ondas de segundo orden*).

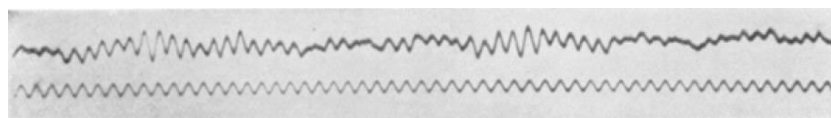


Figura 1. Uno de los primeros registros EEG de Hans Berger. La onda superior es el registro EEG obtenido de un participante, mientras que la onda de la parte inferior representa una onda sinusoidal de 10Hz usada para determinar la frecuencia del registro

Pero ¿cómo es posible que de nuestro cerebro emanen ondas eléctricas? El cerebro es un órgano húmedo por el cual se propaga la electricidad generada por la actividad de las

neuronas. Las neuronas se comunican mediante impulsos nerviosos. No están entrelazadas, sino que entre ellas existen minúsculos espacios de conexión en los que se lleva a cabo la comunicación neuronal. Estos espacios se conocen como sinapsis. Las sinapsis más comunes son las de tipo químico (véase Hammond y Esclapez 2015), donde la neurona emisora (*neurona presináptica*) segrega una serie de compuestos químicos (*neurotransmisores*), que son captados por la neurona receptora (*neurona postsináptica*). Según el tipo de neurotransmisor liberado, la sinapsis puede ser excitatoria o inhibitoria. Si la sinapsis es excitatoria, se desencadena una descarga eléctrica (*potencial de acción*) que recorre la neurona postsináptica y da lugar, a su vez, a la secreción de neurotransmisores en los espacios sinápticos que esta comparte con otras neuronas. Así, la información transmitida por la neurona presináptica a la neurona postsináptica sigue viajando por el tejido nervioso, generando más potenciales de acción en otras neuronas. En cambio, si la sinapsis es inhibitoria, es difícil que se produzca un potencial de acción en la neurona postsináptica, lo cual obstaculizará o impedirá que la neurona postsináptica siga transmitiendo información al resto de neuronas.

Es posible captar la electricidad generada en los potenciales de acción mediante electrodos situados en la superficie del cráneo. Esto convierte a la EEG en un método no invasivo que proporciona una medida directa y en tiempo real de la actividad neuronal. La EEG recoge la actividad eléctrica de la corteza cerebral, esto es, del tejido neuronal situado en la capa más externa del cerebro. En concreto, la actividad eléctrica que captan los electrodos es la suma de la actividad eléctrica excitatoria e inhibitoria de los potenciales de acción de las neuronas piramidales, las cuales destacan por su gran tamaño y forma cónica de su *soma* (el cuerpo de la neurona), además de por ser de las más abundantes en la corteza cerebral y por tener un papel relevante en los procesos cognitivos complejos y en el control motor. La suma de la actividad generada por la acción simultánea y coordinada de grandes conjuntos de neuronas piramidales se refleja en un flujo constante de ondas, de variable amplitud (entre -100 y +100 microvoltios) y frecuencia (desde 0 Hz hasta 40 Hz o más). Cabe destacar que la frecuencia de las ondas se asocia a distintos estados de conciencia (en la Figura 2 se muestran ondas de distinta frecuencia).

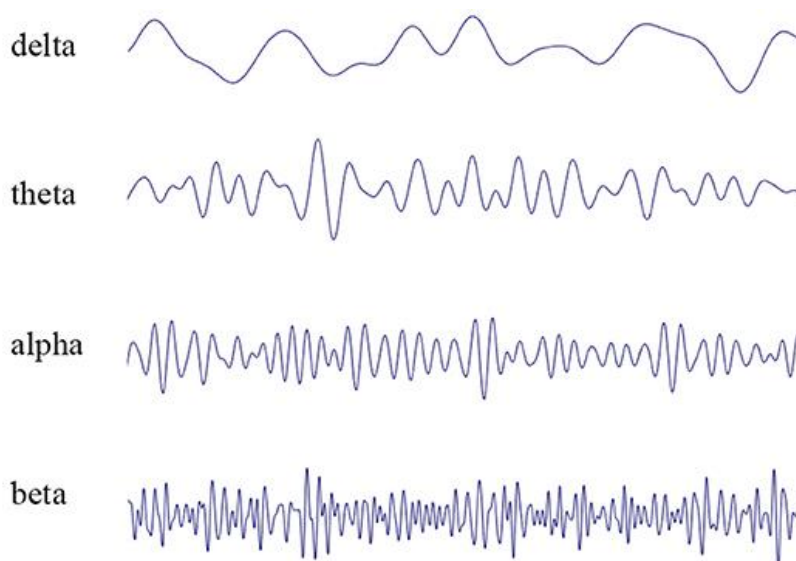


Figura 2. Ondas de diferente frecuencia que se pueden observar en un registro EEG. Aparecen ordenadas de menor a mayor frecuencia (Sun et al. 2020: 2)

Las ondas de primer orden detectadas por Berger (1929), de una frecuencia de 10 Hz, se denominan *ondas alfa* y se asocian a estados de relajación y tranquilidad, y también a fases previas al sueño. Berger también observó unas ondas de frecuencia más elevada, que denominó ondas de segundo orden, con una frecuencia comprendida entre 13 y 30 Hz, las cuales conocemos hoy en día como *ondas beta*. Estas se asocian a estados de actividad mental entre moderada e intensa durante la vigilia. Otras ondas de menor frecuencia fueron observadas por primera vez por Walter (1936) y Walter y Dovey (1944): las *ondas delta* y *theta*. Las ondas theta poseen una frecuencia entre 3,5 y 7,5 Hz y se suelen observar durante fases de sueño ligero. Las ondas delta, por su parte, poseen una frecuencia menor, entre 1 y 3 Hz, y se asocian a fases de sueño profundo.

Aunque existe toda una línea de investigación dedicada a examinar la relación entre el procesamiento del lenguaje y los cambios en la frecuencia de la actividad EEG (véase Petsche y Etlinger 1998, para una revisión), el presente trabajo se centrará en cómo se ha abordado el estudio del procesamiento y representación del lenguaje mediante el análisis de potenciales evocados, el cual será tratado a partir del apartado siguiente.

3. Potenciales evocados

Un potencial evocado (PE; *event-related potential*, ERP, en inglés) se corresponde con la actividad eléctrica, dentro del flujo continuo de la actividad EEG, causada por la presentación de un estímulo (p.ej., una imagen, una palabra, un sonido, etc.). Dicho estímulo *evoca* una serie de cambios en la actividad cerebral, distinguibles de la actividad de fondo, los cuales permiten entrever los mecanismos cerebrales subyacentes al procesamiento del estímulo.

El registro de PE se realiza mediante un conjunto de electrodos colocados sobre la superficie del cráneo (Figura 3). Estos electrodos captan la actividad eléctrica de grandes grupos de neuronas que se activan de forma simultánea y sincronizada durante el procesamiento de un estímulo. En concreto, captan la actividad de las neuronas piramidales de la corteza cerebral, que se encuentran dispuestas de forma perpendicular al lugar donde está situado cada electrodo. Los electrodos se ubican estratégicamente en diferentes zonas del cráneo: frontal, central, parietal, occipital, etc. Esta disposición tiene como finalidad recoger la actividad de diferentes áreas cerebrales.

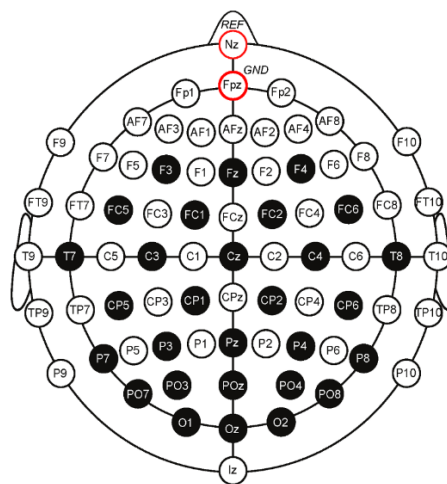


Figura 3. Localización de un conjunto de electrodos siguiendo el sistema internacional 10-20 (Yeom et al. 2014: 2). Se puede encontrar más información sobre el sistema internacional 10-20 en Silverman (1963)

Sin embargo, cabe destacar que la resolución espacial es una de las mayores limitaciones del registro de PE. Es difícil determinar con certeza en qué área cortical se ha generado la actividad captada en un electrodo o electrodos en particular. Por ejemplo, la actividad recogida en un electrodo frontal puede provenir de la corteza frontal, pero también de otras áreas corticales, tanto adyacentes como distantes a esta. Esto se debe a que el cerebro se comporta como un volumen conductor: esto es, la actividad eléctrica generada en una región concreta del cerebro puede propagarse por todo el tejido neuronal. Además, la actividad eléctrica debe superar varios obstáculos hasta lograr ser captada por los electrodos, en concreto, tiene que sobrepasar el resto de tejido cerebral, el cráneo y el cuero cabelludo.

Por todo lo anterior, el registro de PE no es el procedimiento de primera elección para el estudio de la localización espacial de los procesos cognitivos (aunque existen métodos para mejorar su resolución; en esta línea véase Nunez et al. 1994). No obstante, son numerosas las ventajas de esta técnica, que sobresale por su elevada resolución temporal (del orden de milisegundos), lo cual es su principal punto positivo respecto a otras técnicas de exploración neurofisiológica (p.ej., resonancia magnética funcional). También cabe destacar que se trata de una técnica no invasiva, que no requiere una respuesta explícita del participante, y que proporciona una medida multidimensional del procesamiento cognitivo, es decir, aporta más de una variable dependiente para el análisis (véase Luck 2005). Otra importante ventaja es que el registro de PE puede ser descompuesto en diferentes segmentos para su análisis. De esta forma, es posible segmentar una señal de EEG continua en los componentes que la forman, los cuales reflejan distintas fases del procesamiento de un estímulo. Este aspecto se describe con más detalle a continuación.

3.1. Potenciales evocados: componentes

Tras un proceso de limpieza, filtrado y promediado de la actividad EEG, obtenemos una línea que fluctúa en forma de ondas a lo largo de un plano formado por dos ejes (Figura 4). En el eje vertical podemos observar la amplitud de voltaje (en microvoltios), y en el eje horizontal, el tiempo de presentación del estímulo (en milisegundos). El eje horizontal comienza en el milisegundo 0 (o bien unos milisegundos antes, si incluimos la actividad eléctrica de línea base), indicando el momento exacto en el que se ha presentado el estímulo. La actividad eléctrica oscila durante el transcurso temporal, elevándose y hundiéndose. Estas oscilaciones en forma de colinas y valles se denominan *componentes*. La polaridad de los componentes indica si la actividad neuronal durante ese rango temporal es principalmente excitatoria (*polaridad positiva*) o inhibitoria (*polaridad negativa*). Los componentes de polaridad positiva se identifican con la letra *P*, mientras que los de polaridad negativa se identifican con la letra *N*. Es importante tener en cuenta que, al representar la actividad EEG en un gráfico, es práctica habitual invertir la polaridad, de modo que la polaridad negativa queda en la parte superior, mientras que la positiva se encuentra en la parte inferior. Además, la nomenclatura asignada a los componentes refleja el rango temporal o el orden secuencial en el que suelen aparecer. Por ejemplo, el componente N400 es un componente que suele alcanzar su pico de máxima amplitud en torno a los 400 milisegundos tras la presentación del estímulo.

También podemos encontrar componentes numerados como P1, N1, P2, N2, etc., los cuales indican su orden de aparición (p.ej., el componente P2 es el segundo componente de polaridad positiva que podemos observar en el registro). Además, cabe destacar que ciertos componentes poseen una distribución espacial concreta. Es posible que un componente se manifieste en electrodos situados en una región determinada (p.ej., en la región anterior), pero que su magnitud se vea disminuida o que sea indistinguible del resto de fluctuaciones del registro en otras regiones (p.ej., regiones central y parietal en el componente N400; Kutas y Federmeier 2011).

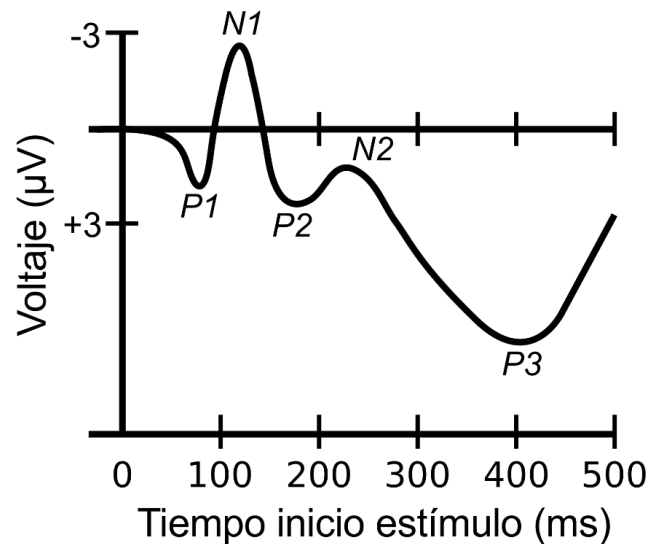


Figura 4. Una representación de los distintos componentes que se pueden apreciar en un registro de PE

Son numerosos los estudios que han investigado los procesos cognitivos asociados a cada uno de los componentes que aparecen en un registro de PE. La lógica detrás de la mayoría de estos estudios es simple. Se crean diversas condiciones experimentales compuestas por estímulos que difieren en alguna variable en concreto, pero igualados en todas aquellas variables ajenas a la manipulación experimental y que pueden influir en los resultados. Los estímulos son presentados a una muestra de participantes, quienes deben realizar alguna tarea con ellos (en ocasiones esto no es necesario, basta con que los participantes sean expuestos a los estímulos). Durante la sesión se registra la actividad EEG de los participantes. Una vez finalizado el experimento, se compara el registro PE entre condiciones para analizar las diferencias de amplitud de los diferentes componentes. Si se observa alguna diferencia entre condiciones en uno o varios componentes, se asume que ese o esos componentes reflejan el proceso cognitivo implicado en el procesamiento de la variable manipulada.

El estudio de Barber et al. (2013) es un buen ejemplo de lo anterior. En este estudio se presentaron 120 palabras y 120 no palabras a un grupo de participantes en una tarea de decisión léxica. Los participantes debían indicar si cada uno de los estímulos presentados se correspondía con una palabra o no. Se manipuló el grado de concreción de las palabras. De las 120 palabras, 60 eran concretas (p.ej., *peló*) y las otras 60 eran abstractas (p.ej., *cosa*). Las palabras de ambos conjuntos estaban igualadas en un gran número de variables: frecuencia léxica, familiaridad, número de letras, número de sílabas, valencia emocional, edad de adquisición, etc. Los resultados mostraron diferencias entre condiciones en distintos componentes. En una franja temporal que abarcaba desde los 250

hasta los 450 milisegundos tras la presentación del estímulo se observó una fluctuación de polaridad negativa en el registro, la cual los autores identificaron con el componente N400. Las palabras concretas obtuvieron una amplitud de voltaje más negativa que las palabras abstractas en la N400. También se observaron diferencias entre condiciones en un componente de polaridad negativa que apareció en torno a los 700 y 1000 milisegundos tras la presentación del estímulo (N700). De nuevo, las palabras concretas mostraron amplitudes de voltaje más negativas que las palabras abstractas en este componente. Teniendo en cuenta que las palabras concretas y abstractas estaban igualadas en un gran número de variables, los autores interpretaron estos resultados como una evidencia de que los componentes N400 y N700 reflejan ciertos mecanismos cognitivos asociados al procesamiento del grado de concreción de las palabras. Según los autores, las diferencias en estos dos componentes indicarían que, durante el procesamiento de una palabra concreta, se activa una mayor cantidad de información semántica que durante el procesamiento de una palabra abstracta. Esto sería debido a que las palabras concretas, a diferencia de las abstractas, requieren la activación e integración de información multimodal (sensoriomotora) representada de forma distribuida en múltiples redes neuronales. En cambio, la representación semántica de las palabras abstractas no incluiría información multimodal, dado que dependería de asociaciones superficiales con otras palabras.

El estudio de Barber et al. (2013) es solo un ejemplo de las múltiples posibilidades que ofrece la técnica de potenciales evocados para el estudio del lenguaje. El uso de esta técnica ha permitido descubrir numerosos componentes vinculados al procesamiento y representación del lenguaje. A continuación, se describen brevemente algunos de los más estudiados y que están más estrechamente relacionados con aspectos lingüísticos.

3.2. Potenciales evocados relacionados con el procesamiento y representación del lenguaje

3.2.1. Potencial de disparidad: Fonología

El potencial de disparidad (*mismatch negativity*, *MMN*, en inglés) es un componente de polaridad negativa, que aparece entre 100 y 200 milisegundos tras la presentación de un estímulo auditivo (véase Figura 5). Este potencial se asocia al procesamiento de estímulos auditivos de cualquier tipo, incluyendo sonidos lingüísticos. Se manifiesta al presentar un sonido inesperado o que difiere en alguna propiedad con respecto a otros sonidos presentados inmediatamente antes. Por ejemplo, es posible observar un MMN si, tras presentar una secuencia de sonidos idénticos, se reproduce de forma inesperada un sonido que se diferencia de los sonidos anteriores en su tono o duración. Esta particularidad del MMN lo convierte en un recurso útil para el estudio de la percepción del habla (véase Näätänen et al. 2007, para una revisión), aunque también se ha usado, en menor medida, para estudiar el acceso léxico y los procesos sintácticos (véase Pulvermüller y Shtyrov 2006, para una revisión). En general, se asume que la modulación de este componente responde a la diferencia entre el sonido que se está presentando al participante y la huella de memoria que ha dejado el sonido anterior.

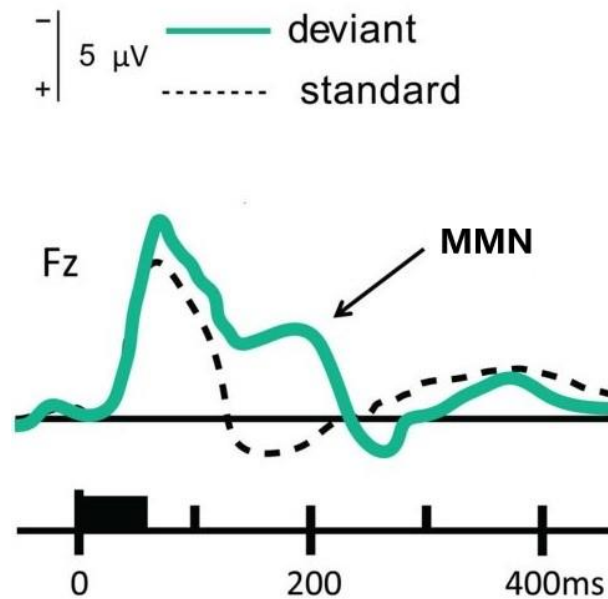


Figura 5. Representación gráfica de un registro del potencial de disparidad (MMN), extraída de Chennu y Bekinschtein (2012: 2). Esta figura y las sucesivas se presentan con la polaridad invertida

El MMN se ha relacionado con la habilidad de los hablantes para discriminar fonemas que pertenecen a la misma o distinta categoría fonológica. Normalmente se observa un MMN de mayor amplitud al discriminar fonemas que pertenecen a distintas categorías fonológicas respecto a fonemas que forman parte de la misma categoría (p.ej., Dehaene-Lambertz 1997). También se han encontrado modulaciones en este componente al discriminar entre fonemas de la lengua nativa y de otras lenguas. Por ejemplo, Näätänen et al. (1997) informaron de un incremento en el componente MMN al exponer a los participantes a un fonema diferente a los presentados en una secuencia de fonemas idénticos, pero este incremento solo se dio cuando el fonema discordante formaba parte del repertorio de fonemas de la lengua nativa del hablante. Por el contrario, no encontraron ningún efecto en el MMN cuando el fonema discordante no pertenecía a su lengua nativa. Otra interesante línea de investigación en la que se ha examinado este componente es en el estudio de la percepción del habla en bebés. Cheour et al. (1998) examinaron a partir de qué edad los bebés pierden la capacidad para discriminar fonemas que no forman parte de su lengua nativa y en qué momento incrementan la capacidad para discriminar los fonemas de su lengua nativa. En los resultados de su estudio se pudo observar un incremento del MMN en bebés de 6 meses al discriminar tanto fonemas de su lengua nativa como de otras lenguas. Sin embargo, cuando se volvió a examinar a esos mismos bebés a la edad de 12 meses, se apreció una reducción significativa de la amplitud del MMN, con respecto a la primera evaluación, cuando se les presentaron fonemas de una lengua diferente de la nativa. En cambio, la amplitud del MMN se incrementó en relación con la primera evaluación al presentarles fonemas de la lengua nativa. Estos resultados sugieren que alrededor de los 12 meses los bebés ven disminuida su capacidad para discriminar sonidos de otras lenguas y, al mismo tiempo, incrementan su capacidad para discriminar sonidos de su lengua materna. Por otra parte, hay evidencia de que la amplitud del MMN refleja procesos de integración de información visual y auditiva durante la percepción de fonemas; por ejemplo, esto se ha observado en tareas donde a los participantes se les presenta un fonema al mismo tiempo que se les facilita información visual sobre la articulación del fonema (Möttönen et al. 2002). Además, se

ha podido comprobar que este componente es sensible a la probabilidad fonotáctica (Bonte et al. 2005), esto es, la probabilidad de que un sonido o secuencia de sonidos esté presente en un determinado idioma. En resumen, tanto los resultados obtenidos con bebés como los registrados en adultos muestran que el MMN es un recurso de gran utilidad para examinar el desarrollo de las representaciones de sonidos lingüísticos, así como los procesos de percepción y producción del habla.

3.2.2. N400: Semántica

El componente N400 es posiblemente uno de los más estudiados en relación con el procesamiento y representación del lenguaje. Se trata de un componente de polaridad negativa, de distribución centro-parietal, que aparece entre los 200 y 600 milisegundos y que alcanza su pico de mayor amplitud en torno a los 400 milisegundos tras la presentación de estímulos visuales y auditivos (véase Figura 6). Se asocia, sobre todo, al procesamiento semántico de estímulos lingüísticos de diversa complejidad, por ejemplo, palabras, enunciados y discursos (véase Kutas y Federmeier 2011, para una revisión), aunque también se han observado efectos en la N400 asociados al procesamiento morfosintáctico, como es el caso de la concordancia verbal (Mancini et al. 2011). Este componente fue descubierto por Kutas y Hillyard en 1980. En este estudio pionero examinaron la actividad eléctrica cerebral al presentar al final de una oración una palabra, cuyo grado de relación con el contexto generado por la oración variaba. Algunas palabras estaban ligeramente relacionadas con el contexto: *He took a sip from the waterfall* (Él bebió un sorbo de la cascada), mientras que otras no estaban relacionadas: *He took a sip from the transmitter* (Él bebió un sorbo del transmisor). La actividad eléctrica generada por estas palabras fue comparada con la de palabras apropiadas al contexto, por ejemplo, *It was his first day at work* (Era su primer día de trabajo). Los resultados mostraron un aumento de la negatividad del componente N400 durante la lectura de palabras inapropiadas al contexto con respecto a palabras apropiadas al contexto. Además, este incremento fue mayor para las palabras nada relacionadas respecto a las ligeramente relacionadas.

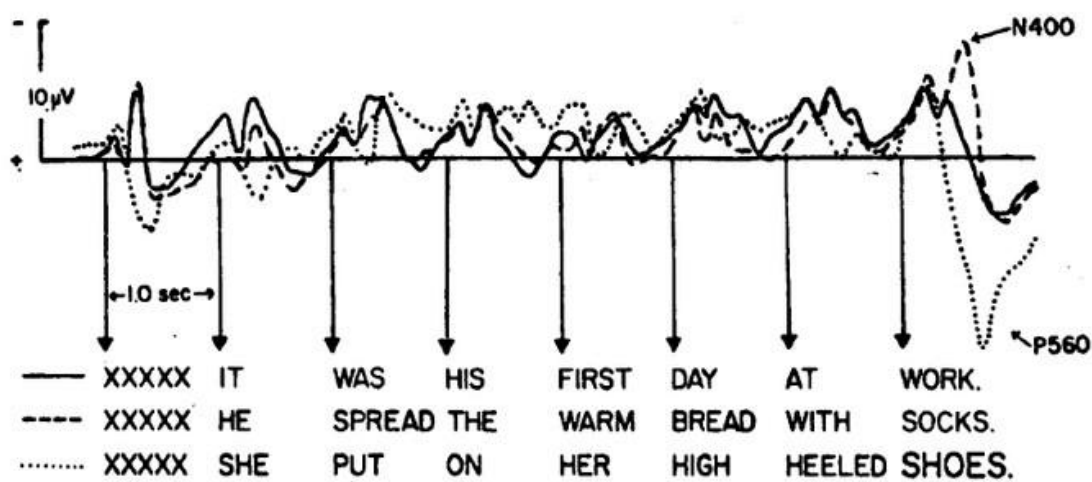


Figura 6. Representación gráfica de un registro del componente N400 (Kutas y Hillyard 1980: 203)

Existen, al menos, dos posibles interpretaciones de este efecto. Una de ellas sugiere que, durante la lectura de un enunciado, el lector va formulando hipótesis sobre la palabra que aparecerá a continuación, en un intento de predecir y lograr integrar dicha palabra en el contexto generado por el enunciado. De este modo, la lectura se entiende como un proceso incremental de acumulación de evidencias y, por tanto, las diferencias observadas en la N400 entre palabras apropiadas e inapropiadas al contexto se deben a que el significado de la palabra inapropiada no se ajusta a las hipótesis generadas por el lector. En este sentido, el incremento en el componente N400 ante palabras incongruentes al contexto reflejaría una disonancia entre la predicción del lector y la palabra presentada, así como los intentos de integrar esa palabra en el contexto semántico de la frase (Brown y Hagoort 1993). Otra hipótesis sugiere que las diferencias en la N400 reflejan la cantidad de preactivación semántica de la palabra (Kutas et al. 1984). Según esta hipótesis, la representación semántica de la palabra va recibiendo activación de forma inconsciente y automática durante la lectura de las palabras previas. Cuanto mayor es la relación entre la palabra y las palabras previas, más activación recibirá dicha palabra y menor será la cantidad de activación necesaria para integrarla en el contexto. Así, las diferencias en la N400 indicarían que las palabras incongruentes, al no haber recibido apenas activación del resto de palabras que componen el enunciado, requerirán un mayor esfuerzo de procesamiento para lograr una activación suficiente que permita su comprensión e integración en el contexto. En cambio, la preactivación de las palabras congruentes será más alta, por lo que hará falta menos activación para lograr integrarlas.

También se han observado efectos de la N400 durante el procesamiento de palabras precedidas por un contexto mínimo, esto es, por el contexto generado por una palabra presentada inmediatamente antes. Por ejemplo, se han examinado los efectos en la N400 mediante el paradigma de priming semántico. Este paradigma consiste en presentar un estímulo (*prime*) que mantiene algún tipo de relación semántica con un estímulo que se presenta a continuación y al que los participantes deben responder (*target*). El *prime* puede presentarse o bien subliminalmente (muy brevemente, con una duración de apenas 50 milisegundos o menos, impidiendo que el participante logre percibirlo conscientemente) o bien de forma visible. Si el paradigma de *priming* se utiliza, por ejemplo, con una tarea de decisión léxica, el participante debe indicar si el *target* (p.ej., *gato*) es una palabra o no. Dicho *target* puede ir precedido por un *prime* relacionado semánticamente (p.ej., *perro*) o por un *prime* no relacionado semánticamente (p.ej., *silla*). De esta manera, se mide la latencia de respuesta en la condición congruente (p.ej., *perro-gato*) y esta se compara con la latencia de respuesta de la condición incongruente (p.ej., *silla-gato*). Así, se asume que cualquier diferencia entre condiciones en la latencia de respuesta refleja la influencia que ejerce el *prime* sobre el *target*. Usando este paradigma con la tarea de decisión léxica se ha observado que los *targets* precedidos por *primes* semánticamente relacionados muestran una N400 de menor amplitud que los *targets* precedidos por *primes* no relacionados (p.ej., Holcomb 1993). Esto sucede incluso cuando los *primes* son presentados subliminalmente y, por tanto, no son percibidos y procesados de forma consciente por los participantes (p.ej., Kiefer 2002). Toda esta evidencia sugiere que las diferencias en la N400 reflejan el grado de preactivación que recibe la representación semántica del *target* por parte del *prime*. A mayor relación semántica entre *prime* y *target*, mayor activación recibe la representación semántica del *target* antes incluso de que sea presentado y, en consecuencia, menor cantidad de información semántica del *target* será necesaria para poder reconocerlo.

Numerosos estudios sugieren que la amplitud de la N400 es también una medida de la calidad y cantidad de la información semántica de una palabra, lo que la convierte en una valiosa herramienta para examinar la representación y acceso al significado de las palabras. Esto se ha podido comprobar, sobre todo, en estudios donde se ha examinado el procesamiento de palabras aisladas (esto es, sin contexto previo) mediante la tarea de decisión léxica. Por ejemplo, existe evidencia de que (i) las palabras concretas generan amplitudes N400 mayores que las palabras abstractas (Kounios y Holcomb 1994); (ii) las palabras con mayor cantidad de rasgos semánticos se asocian con amplitudes N400 más elevadas que las palabras con pocos rasgos semánticos (Rabovsky et al. 2012); (iii) las palabras con un gran número de asociados léxicos (palabras que suelen concurrir juntas; p.ej., *gato-negro*) muestran una N400 mayor que las palabras con pocos asociados léxicos (Laszlo y Federmeier 2011), y (iv) las palabras ambiguas muestran una N400 mayor que las no ambiguas (Haro et al. 2017).

En definitiva, existe numerosa evidencia que muestra la utilidad del componente N400 para el análisis del procesamiento semántico, tanto de palabras embebidas en contextos como en palabras aisladas, así como para estudiar los efectos de la riqueza semántica en el procesamiento léxico.

3.2.3. P600: Sintaxis

El componente P600 es un componente de polaridad positiva, de distribución posterior, que suele emerger alrededor de 500 milisegundos después del inicio de la presentación de un estímulo lingüístico y cuya duración es de aproximadamente 200 milisegundos (véase Figura 7). Además, su pico de mayor amplitud se manifiesta en torno a los 600 milisegundos.

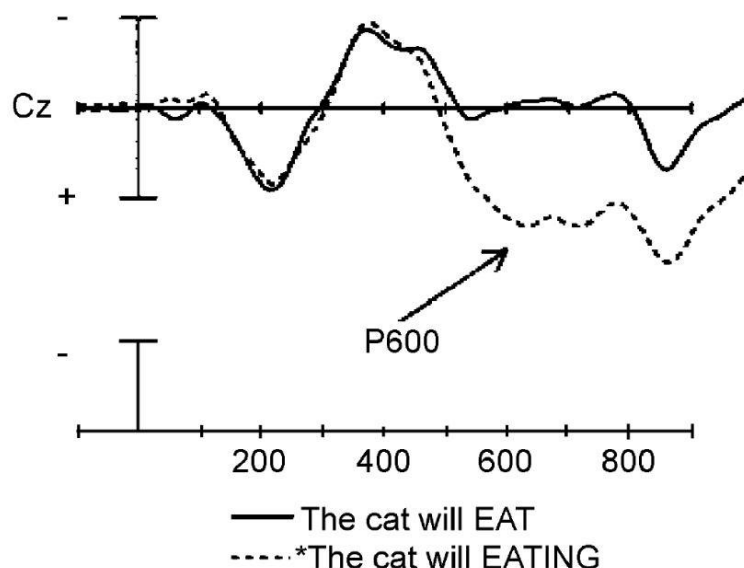


Figura 7. Representación gráfica de un registro del componente P600 (Osterhout et al. 2008: 511)

El descubrimiento de este componente se remonta al estudio de Osterhout y Holcomb (1992). Este estudio ejemplifica el tipo de efectos que se pueden observar en el componente P600 y las manipulaciones experimentales que dan lugar a dichos efectos. En el primer experimento del estudio, los autores presentaron frases como las siguientes:

- (1) The woman struggled to prepare the meal
'La mujer se esforzó para preparar la comida'
- (2) The woman persuaded to answer the door
'La mujer persuadió para contestar la puerta'

Como se puede observar, la frase (1) es gramaticalmente correcta, mientras que la (2) no lo es. Dado que *persuade* es un verbo transitivo, requiere un objeto directo para dar sentido a la oración. En tal caso, al leer la preposición *to* de la oración (2), el analizador sintáctico se encontraría ante un incumplimiento de las propiedades de subcategorización del verbo *persuade*, algo que no ocurriría en la oración (1). Esta incongruencia sintáctica se vio reflejada en el registro de PE de Osterhout y Holcomb (1992). En concreto, se observó un incremento de voltaje de polaridad positiva entre los 500 y 800 milisegundos tras leer la preposición *to* en oraciones agramaticales como la presentada en el ejemplo (2) en comparación con oraciones gramaticales como la del ejemplo (1).

El estudio de Osterhout y Holcomb (1992) sugiere que el componente P600 es un indicador de las dificultades del analizador sintáctico al lidiar con ciertas anomalías o incongruencias sintácticas. En general, los efectos en la P600 se manifiestan cuando la continuación sintáctica de una frase es poco o nada plausible gramaticalmente. Se han observado efectos en el componente P600 ante una gran diversidad de anomalías sintácticas: errores de concordancia entre el verbo y el sujeto de una oración (p.ej., Hagoort et al. 1993), anomalías en la estructura de una oración (Hahne y Friederici 1999), errores de concordancia de caso (Münte et al. 1998), de género (Gunter et al. 2000) y de número (Münte et al. 1997) y errores de inflexión del verbo (Gunter et al. 1997). Además, se ha encontrado una modulación de la P600 en frases gramaticalmente correctas, pero cuya estructura sintáctica es compleja o bien no refleja la opción sintáctica preferida por los hablantes (Kaan et al. 2000), así como en oraciones que presentan una ambigüedad sintáctica temporal, en las cuales existen varios análisis sintácticos correctos durante su procesamiento (Kaan y Swaab 2003). Por último, también cabe destacar que algunos estudios han observado efectos en la P600 no relacionados estrictamente con el procesamiento sintáctico (para profundizar más véase Bornkessel-Schlesewsky y Schlewsky 2008).

4. Cómo llevar a cabo un estudio de potenciales evocados

Para finalizar este trabajo, se ilustra el proceso de diseño, administración y análisis de un experimento de potenciales evocados. Para ello se describirá el procedimiento que seguimos en un estudio realizado en nuestro laboratorio (Haro et al. 2017), en el cual se comparó el procesamiento de palabras ambiguas (p.ej., *banco*) y de palabras no ambiguas (p.ej., *sandía*). Cabe puntualizar que los pasos que se describen a continuación no abordan de forma exhaustiva todo el procedimiento, ni tampoco pretenden mostrar la forma ideal de llevarlo a cabo; es simplemente una breve guía que recoge algunas recomendaciones para aquellos investigadores que se enfrentan por primera vez a la implementación y análisis de un experimento de potenciales evocados.

Existe evidencia de que las palabras ambiguas se reconocen más rápido que las no ambiguas en tareas de decisión léxica (p.ej., Pexman et al. 2004). Se ha propuesto que esta facilitación se debe a que las palabras ambiguas, al poseer más de una representación

semántica, generarían una mayor cantidad de activación semántica que las no ambiguas durante su procesamiento (p.ej., Borowsky y Masson 1996; Hino y Lupker 1996). Dicho incremento en la activación de las palabras ambiguas aumentaría la familiaridad de estas palabras, lo que en última instancia aceleraría su reconocimiento. Para poner a prueba esta hipótesis, se examinó el componente N400 durante una tarea de decisión léxica. Como se ha expuesto anteriormente, hay indicios de que el componente N400 se ve modulado por la cantidad y calidad de la información semántica. Esto es, cuanto mayor es la cantidad o calidad de la información semántica, más elevada es la amplitud de la N400 (p.ej., Laszlo y Federmeier 2011). Por tanto, si la facilitación de las palabras ambiguas se debe a que estas activan una mayor cantidad de información semántica durante su procesamiento, sería de esperar que generasen una amplitud de la N400 superior a la de las palabras no ambiguas.

Además de lo anterior, también se analizó si la relación entre los significados de las palabras ambiguas influye en su procesamiento. Existen palabras ambiguas de significados relacionados (esto es, polisémicas; p.ej., *aguja*) y palabras ambiguas de significados no relacionados (esto es, homónimas; p.ej., *heroína*). Algunos estudios han mostrado que las palabras homónimas se reconocen más lentamente que las polisémicas (p.ej., Rodd et al. 2002). Esta inhibición en el reconocimiento de las palabras homónimas se ha interpretado como una evidencia de que los significados no relacionados de estas palabras están representados de forma separada en la memoria y que compiten durante el procesamiento (Rodd et al. 2004). Para poner a prueba esta hipótesis, se examinó la amplitud del componente N400 durante el reconocimiento de palabras polisémicas y homónimas. De existir diferencias representacionales o de procesamiento entre estas palabras, sería esperable observar un patrón de resultados distinto entre ellas en el componente N400 durante la tarea de decisión léxica. En concreto, en línea con algunos estudios previos (p.ej., Beretta et al. 2005), se esperaría una mayor amplitud de la N400 en palabras homónimas respecto a palabras polisémicas.

Las variables de interés en este experimento fueron, por tanto, el número de significados y la relación entre los significados (ambas basadas en los juicios subjetivos de los participantes). De modo que se seleccionó un conjunto de palabras que diferían en estas variables. En total se usaron 76 palabras ambiguas y 76 no ambiguas. De las 76 ambiguas, la mitad eran homónimas y la otra mitad eran polisémicas. En este punto hay que hacer un breve inciso. Es muy importante que el número de estímulos por condición en experimentos de PE sea elevado. Se recomienda un mínimo de 30 estímulos por condición. La razón de esto es que el registro EEG de donde se extraen los componentes PE es muy ruidoso, es decir, gran parte de la actividad eléctrica recogida en este registro no tiene relación con el proceso que se está examinando. Para reducir todo este ruido y poder extraer los componentes PE, es necesario obtener un gran número de registros de cada uno de los estímulos. De esta forma, se logra mejorar la proporción entre la señal (la actividad EEG que nos interesa) y el ruido (la actividad EEG que no nos interesa). Además, dado que el objetivo del estudio era evaluar únicamente el efecto del número de significados y de la relación entre significados, se igualaron los conjuntos de palabras en todas aquellas variables que podían tener un efecto en el registro de PE: por ejemplo, frecuencia léxica, número de letras, grado de concreción, edad de adquisición, etc. De no haber igualado estas variables entre los conjuntos, es posible que el registro de PE se hubiera visto “contaminado” por su efecto.

Pasemos ahora al procedimiento que debemos seguir para administrar el experimento. Un estudio como este requiere (i) un ordenador para la presentación de los estímulos del experimento y el registro de las respuestas conductuales (tiempos de respuesta y errores); (ii) otro ordenador para el registro EEG; (iii) un conjunto de electrodos y un gorro donde irán acoplados estos electrodos (en Haro et al. 2017 se usó un *Acticap* de 32 electrodos de la marca *Brain Products*), y (iv) un amplificador de la señal EEG (en Haro et al. 2017 se usó un *Actichamp*, también de *Brain Products*), al cual se debe conectar tanto el conjunto de electrodos como el ordenador de registro EEG. Por supuesto, también será necesario un enlace de conexión entre el ordenador de presentación de estímulos y el de registro EEG, para así poder identificar los segmentos del registro EEG que se corresponden con la presentación de cada estímulo. Además, se requiere un par de electrodos para registrar artefactos oculares. Los artefactos oculares se producen cuando el participante realiza movimientos horizontales o verticales de los ojos que alteran el registro de EEG. Por ejemplo, un parpadeo produce una alteración en el registro de EEG que afecta al registro de bastantes electrodos. Por último, para reducir la resistencia eléctrica de los electrodos, es preciso disponer de un gel conductor, que se administra sobre el cuero cabelludo mediante una jeringa.

Si se dispone de todo lo anterior, se puede proceder a la preparación del participante. El participante ha de sentarse en un asiento cómodo, donde permanecerá durante toda la sesión experimental. Se le coloca el gorro con electrodos sobre la cabeza (en Haro et al. 2017 se empleó un gorro de 32 electrodos), asegurando que el electrodo central (CZ) se sitúa justo en el centro de la parte superior de su cráneo, en línea con sus orejas. Se verifica que el gorro está bien ajustado a su cráneo, sin arrugas y firmemente sujeto. Ahora se conecta el electrodo de toma de tierra, el cual se suele colocar en la parte frontal y se utiliza para reducir las interferencias producidas por el ruido de la red eléctrica, y los electrodos o electrodo de referencia. Un lugar habitual donde colocar el electrodo de referencia es el lóbulo de la oreja o bien el hueso mastoide, situado detrás de la oreja. Entonces se comienza a aplicar el gel mediante la jeringa en cada uno de los electrodos. Los gorros disponen de una entrada en cada electrodo para facilitar la administración del gel. Es muy recomendable comenzar por el electrodo de toma de tierra y los electrodos de referencia. Durante el proceso de administración del gel es importante tener conectado el ordenador de registro, el cual indica la impedancia de cada electrodo, esto es, la resistencia eléctrica que existe entre el electrodo y el cuero cabelludo. Para asegurar un buen registro, es necesario conseguir que la impedancia de los electrodos sea inferior a 5 k Ω . Una vez se ha logrado que la resistencia de los electrodos se sitúe por debajo de ese umbral, se colocan los electrodos que se encargarán de registrar los artefactos oculares. Un lugar habitual para colocarlos es en la mejilla izquierda y en la sien derecha. También se debe reducir la impedancia de estos electrodos.

Tras completar el paso anterior, es momento de poner en marcha el experimento (en el ordenador de presentación de estímulos) y de iniciar el registro EEG. Es importante verificar que la actividad de cada electrodo es normal, esto es, que no hay una cantidad anómala de ruido, que la amplitud de todos ellos es similar y que su línea de registro no se desvía de su canal. Si aparece alguna anomalía, se deberá comprobar que los electrodos de referencia y de toma de tierra funcionan correctamente y que la impedancia de todos los electrodos está por debajo del umbral recomendado (5 k Ω).

Tras obtener el registro EEG de cada participante, se debe realizar una limpieza de datos. En primer lugar, hay que seguir una serie de criterios para eliminar datos EEG en función

de la respuesta conductual de los participantes. En concreto, en Haro et al. (2017) se eliminaron los datos EEG de los participantes con más de un 15 % de errores en la tarea de decisión léxica y también los tiempos de respuesta de cada participante que excedían dos desviaciones típicas de su media de tiempos de respuesta. También se eliminaron de los datos EEG los registros correspondientes a ensayos de respuesta incorrecta o de no respuesta. Para el resto de los pasos se puede usar el software *BrainVision Analyzer 2* (de la marca *Brain Products*). Primero se especifica la actividad eléctrica de referencia a partir de la cual se calculará la actividad de cada canal de registro (p.ej., el promedio de la actividad de los dos electrodos de referencia). Después se aplica un filtro de frecuencias para eliminar las frecuencias excesivamente altas o bajas del registro EEG, por ejemplo, se puede utilizar un *zerophase shift digital filter* de 0,1 a 30 Hz. Seguidamente se especifica el periodo de actividad eléctrica que servirá como línea base, a partir del cual se recalculará la actividad eléctrica de cada ensayo (100 milisegundos en Haro et al. 2017), y después se delimita el rango temporal de actividad eléctrica que se va a analizar en cada ensayo (empezando desde el inicio de la presentación del estímulo). El rango temporal debe ser lo suficientemente amplio para que incluya los componentes que se desea analizar, pero evitando que se solape con la actividad eléctrica del ensayo siguiente (p.ej., un periodo de 800 milisegundos desde el inicio de la presentación del estímulo). A continuación, se deben eliminar las posibles anomalías del registro. Estas comprenden (i) artefactos producidos por movimientos bruscos de los ojos (p.ej., parpadeos); (ii) amplitudes de voltaje excesivamente altas (mayores a 75 microvoltios o inferiores a -75 microvoltios en Haro et al. 2017) y (iii) variaciones de voltaje excesivamente elevadas (mayores a 150 microvoltios o -150 microvoltios en Haro et al. 2017). Tras esto, se seleccionan los segmentos de cada condición experimental y se calcula el promedio de actividad eléctrica de cada una de ellas. Además, hay que seleccionar el rango temporal donde se encuentra el componente que queremos analizar. Por ejemplo, en Haro et al. (2017) se seleccionó el rango comprendido entre 350 y 450 milisegundos, donde se visualizaba una modulación del registro compatible con el componente N400. Por último, se vuelcan los datos en un archivo, el cual debe incluir el promedio de voltaje de cada electrodo en el rango temporal seleccionado.

Después de realizar todo el proceso de limpieza, filtrado y volcado de datos, se puede proceder a analizar los datos obtenidos. Se abre el archivo donde se han volcado los datos con el software que se usará para el análisis de datos. Dependiendo de los objetivos del estudio o bien de las particularidades o distribución de cada componente, se puede seleccionar un conjunto concreto de electrodos a analizar o bien seleccionarlos todos. En Haro et al. (2017) se analizaron todos los electrodos en diferentes análisis de la varianza (ANOVA). En concreto, se realizaron dos ANOVAs de medidas repetidas, en los cuales se incluyeron diversos factores intra-sujetos. El primer ANOVA tenía como objetivo examinar las diferencias entre palabras ambiguas y no ambiguas. En este se incluyeron los factores ambigüedad (palabras ambiguas y no ambiguas) y electrodo (28 electrodos). El análisis reveló un efecto principal del factor ambigüedad, $F(1, 23) = 5,07, p = 0,034$. Las palabras ambiguas mostraron una mayor amplitud en la N400 (-1,70 microvoltios) que las palabras no ambiguas (-1,20 microvoltios). No se halló interacción entre los factores ambigüedad y electrodo, $F(27, 621) = 1,53, p = 0,20$. La interpretación que se dio a este resultado fue que, dado que la amplitud de la N400 es un indicador del grado de activación semántica de las palabras, las palabras ambiguas lograron activar una mayor cantidad de información semántica que las no ambiguas, facilitando esto su reconocimiento en la tarea de decisión léxica. Por otra parte, en el segundo ANOVA se

examinaron las diferencias entre palabras polisémicas y homónimas. En el ANOVA se introdujeron los factores tipo de ambigüedad (palabras polisémicas y palabras homónimas) y electrodo (28 electrodos). No se observó un efecto del factor tipo de ambigüedad, $F(1, 23) = 0,16, p = 0,70$. Las palabras homónimas y polisémicas mostraron amplitudes similares en la N400 (-1,75 y -1,62 microvoltios, respectivamente). Tampoco se observó interacción entre los factores tipo de ambigüedad y electrodo, $F(27, 621) = 1,41, p = 0,24$. Este resultado sugeriría que la activación semántica generada por las palabras polisémicas y homónimas es similar.

Como se puede observar, llevar a cabo un experimento de potenciales evocados, aunque entraña algo más dificultad que un experimento conductual, no es una tarea excesivamente compleja. Si se está interesado en profundizar más, recomendamos consultar el trabajo de Luck (2005), donde se abordan con más detalle todos los pasos involucrados en la planificación, administración y análisis de un experimento de potenciales evocados.

5. Conclusiones

En este trabajo se ha presentado una breve introducción al uso de los potenciales evocados en el estudio del lenguaje. En resumen, la principal ventaja del registro de PE es su resolución temporal, dado que permite realizar un seguimiento milisegundo a milisegundo del procesamiento de estímulos lingüísticos. Se ha demostrado su utilidad para examinar diferentes niveles de procesamiento y representación del lenguaje, por ejemplo, la fonología, la semántica y la sintaxis. Además, permite arrojar luz sobre algunas cuestiones en las que las medidas conductuales muestran resultados inconsistentes. Esto la convierte en una medida que, lejos de reemplazar a las medidas conductuales, puede ser de gran utilidad para complementar los datos obtenidos mediante aquellas. De hecho, el registro de PE también se ha utilizado en combinación con otro tipo de medidas, por ejemplo, de seguimiento de movimientos oculares (p.ej., Henderson et al. 2013; Álvarez García 2022), la magnetoencefalografía (Wang et al. 2020) o la resonancia magnética funcional (Geukes et al. 2013; Marrero-Aguar 2022). Pero no todo son ventajas. Su principal limitación es su baja resolución espacial, por lo que no es la medida más recomendable si se está interesado en examinar la localización cerebral de los procesos cognitivos. Por otra parte, la preparación, administración y análisis de un experimento de potenciales evocados no requiere demasiados conocimientos técnicos. Como se ilustra en el ejemplo presentado anteriormente, un experimento de este tipo se puede llevar a cabo sin demasiados problemas si seguimos una serie de pasos en el orden adecuado.

6. Referencias

- Álvarez García, Esther. 2022. Lo que esconden tus ojos: la metodología eye-tracking aplicada al estudio del lenguaje. *Estudios de Lingüística del Español* 45: 205-239.
- Balota, David A.; Abrams, Richard A. 1995. Mental chronometry: beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21.5: 1289-1302.

- Barber, Horacio A.; Otten, Leun J.; Kousta, Stavroula-Thaleia; Vigliocco, Gabriella. 2013. Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and language* 125.1: 47-53.
- Beretta, Alan; Fiorentino, Robert; Poeppel, David. 2005. The effects of homonymy and polysemy on lexical access: An MEG study. *Cognitive Brain Research* 24.1: 57-65.
- Berger, Hans. 1929. Über das elektroenkephalogramm des menschen. *Archiv für psychiatrie und nervenkrankheiten* 87.1: 527-570.
- Bonte, Milene L.; Mitterer, Holger; Zellagui, Nadia; Poelmans, Hanne; Blomert, Leo. 2005. Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology* 116.12: 2765-2774.
- Bornkessel-Schlesewsky, Ina; Schlesewsky, Matthias. 2008. An alternative perspective on “semantic P600” effects in language comprehension. *Brain research reviews* 59.1: 55-73.
- Borowsky, Ron; Masson, Michael E. 1996. Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22.1: 63-85.
- Brown, Colin; Hagoort, Peter. 1993. The processing nature of the N400: Evidence from masked priming. *Journal of cognitive neuroscience* 5.1: 34-44.
- Chennu, Srivas; Bekinschtein, Tristan A. 2012. Arousal modulates auditory attention and awareness: insights from sleep, sedation, and disorders of consciousness. *Frontiers in psychology* 3: 65.
- Cheour, Marie; Ceponiene, Rita; Lehtokoski, Anne; Luuk, Aavo; Allik, Jüri; Alho, Kimmo; Näätänen, Risto. 1998. Development of language-specific phoneme representations in the infant brain. *Nature neuroscience* 1.5: 351-353.
- Dehaene-Lambertz, Ghislaine. 1997. Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport* 8.4: 919-924.
- Garrod, Simon. 2006. Psycholinguistic research methods. En K. Brown, ed. *Encyclopedia of Language and Linguistics*. Amsterdam: Elsevier, pp. 251-257.
- Geukes, Sebastian; Huster, René J.; Wollbrink, Andreas; Junghöfer, Markus; Zwitserlood, Pienie; Döbel, Christian. 2013. A large N400 but no BOLD effect—comparing source activations of semantic priming in simultaneous EEG-fMRI. *PloS one* 8.12: e84029.
- Gunter, Thomas C.; Friederici, Angela D.; Schriefers, Herbert. 2000. Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of cognitive neuroscience* 12.4: 556-568.
- Gunter, Thomas C.; Stowe, Laurie A.; Mulder, Gusbertus. 1997. When syntax meets semantics. *Psychophysiology* 34.6: 660-676.
- Hagoort, Peter; Brown, Colin; Groothusen, Jolanda. 1993. The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and cognitive processes* 8.4: 439-483.
- Hahne, Anja; Friederici, Angela D. 1999. Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of cognitive neuroscience* 11.2: 194-205.
- Hammond, Constance; Esclapez, Monique. 2015. The chemical synapses. En C. Hammond, ed. *Cellular and Molecular Neurophysiology*. Amsterdam: Academic Press, pp. 121-144.

- Haro, Juan; Demestre, Josep; Boada, Roger; Ferré, Pilar. 2017. ERP and behavioral effects of semantic ambiguity in a lexical decision task. *Journal of Neurolinguistics* 44: 190-202.
- Henderson, John M.; Luke, Steven G.; Schmidt, Joseph; Richards, John E. 2013. Co-registration of eye movements and event-related potentials in connected-text paragraph reading. *Frontiers in systems neuroscience* 7: 28.
- Hino, Yasushi; Lupker, Stephen J. 1996. Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance* 22.6: 1331-1356.
- Holcomb, Phillip J. 1993. Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology* 30.1: 47-61.
- Igoa, José Manuel. Las tareas conductuales en la investigación sobre el procesamiento del lenguaje. *Estudios de Lingüística del Español* 45: 133-158.
- Kaan, Edith; Harris, Anthony; Gibson, Edward; Holcomb, Phillip. 2000. The P600 as an index of syntactic integration difficulty. *Language and cognitive processes* 15.2: 159-201.
- Kaan, Edith; Swaab, Tamara Y. 2003. Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of cognitive neuroscience* 15.1: 98-110.
- Kiefer, Markus. 2002. The N400 is modulated by unconsciously perceived masked words: Further evidence for an automatic spreading activation account of N400 priming effects. *Cognitive Brain Research* 13.1: 27-39.
- Kounios, John; Holcomb, Phillip J. 1994. Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20.4: 804-823.
- Kutas, Marta; Federmeier, Kara D. 2011. Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology* 62: 621-647.
- Kutas, Marta; Hillyard, Steven A. 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207.4427: 203-205.
- Kutas, Marta; Lindamood, Timothy E.; Hillyard, Steven A. 1984. Word expectancy and event-related brain potentials during sentence processing. En E.S. Kornblum y J. Requin, eds. *Preparatory states and processes*. Hillsdale: Lawrence Erlbaum Associates Inc, pp. 217-237.
- Laszlo, Sarah; Federmeier, Kara D. 2011. The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology* 48.2: 176-186.
- Luck, Steven J. 2005. *An introduction to the event-related potential technique*. Cambridge: MIT Press
- Mancini, Simona; Molinaro, Nicola; Rizzi, Luigi; Carreiras, Manuel. 2011. A person is not a number: Discourse involvement in subject-verb agreement computation. *Brain research* 1410: 64-76.
- Marrero-Aguiar, Victoria. 2022. Aportación de las investigaciones con neuroimagen funcional a los estudios lingüísticos: algunas preguntas y respuestas. *Estudios de Lingüística del Español* 45: 159-183.
- Möttönen, Riikka; Krause, Christina M.; Tiippana, Kaisa; Sams, Mikko. 2002. Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Research* 13.3: 417-425.

- Münter, Thomas F.; Heinze, Hans-Jochen; Matzke, Mike; Wieringa, Bernardina M.; Johannes, Sönke. 1998. Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia* 36.3: 217-226.
- Münter, Thomas F.; Szentkuti, Andras; Wieringa, Bernardina M.; Matzke, Mike; Johannes, Sönke. 1997. Human brain potentials to reading syntactic errors in sentences of different complexity. *Neuroscience Letters* 235.3: 105-108.
- Näätänen, Risto; Lehtokoski, Anne; Lennes, Mietta; Cheour, Marie; Huotilainen, Minna; Iivonen, Antti; Vainio, Martti; Alku, Paavo; Ilmoniemi, Risto J.; Luuk, Aavo; Allik, Jüri; Sinkkonen, Janne; Alho, Kimmo. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385.6615: 432-434.
- Näätänen, Risto; Paavilainen, Petri; Rinne, Teemu; Alho, Kimmo. 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical neurophysiology* 118.12: 2544-2590.
- Nunez, Paul L.; Silberstein, Richard B.; Cadusch, Peter J.; Wijesinghe, Ranjith S.; Westdorp, Andrew F.; Srinivasan, Ramesh. 1994. A theoretical and experimental study of high resolution EEG based on surface Laplacians and cortical imaging. *Electroencephalography and clinical neurophysiology* 90.1: 40-57.
- Osterhout, Lee; Holcomb, Phillip J. 1992. Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language* 31.6: 785-806.
- Osterhout, Lee; Poliakov, Andrew; Inoue, Kayo; McLaughlin, Judith, Valentine, Geoffrey; Pitkanen, Ilona; Frenck-Mestre, Cheryl; Hirschensohn, Julia. 2008. Second-language learning and changes in the brain. *Journal of neurolinguistics* 21.6: 509-521.
- Petsche, Hellmuth; Etlinger, Susan C. 1998. *EEG and thinking - power and coherence analysis of cognitive processes*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- Pexman, Penny M.; Hino, Yasushi; Lupker, Stephen J. 2004. Semantic ambiguity and the process of generating meaning from print. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30.6: 1252-1270.
- Pulvermüller, Friedemann; Shtyrov, Yury. 2006. Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Progress in neurobiology* 79.1: 49-71.
- Rabovsky, Milena; Sommer, Werner; Rahman, Rasha Abdel. 2012. Depth of conceptual knowledge modulates visual processes during word reading. *Journal of Cognitive Neuroscience* 24.4: 990-1005.
- Rodd, Jennifer; Gaskell, Gareth; Marslen-Wilson, William. 2002. Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language* 46.2: 245-266.
- Rodd, Jennifer.; Gaskell, Gareth; Marslen-Wilson, William. 2004. Modelling the effects of semantic ambiguity in word recognition. *Cognitive science* 28.1: 89-104.
- Silverman, Daniel. 1963. The rationale and history of the 10-20 system of the International Federation. *American Journal of EEG Technology* 3.1: 17-22.
- Sun, Yi; Wei, Changwei; Cui, Victoria; Xiu, Meihong; Wu, Anshi. 2020. Electroencephalography: Clinical applications during the perioperative period. *Frontiers in Medicine* 7: 251.

- Walter, W. Grey. 1936. The location of cerebral tumours by electro-encephalography. *The Lancet* 2: 305-308.
- Walter, William Grey; Dovey, Vivian Joan. 1944. Electro-encephalography in cases of sub-cortical tumour. *Journal of neurology, neurosurgery, and psychiatry* 7.3-4: 57-65.
- Wang, Lin; Wlotko, Edward; Alexander, Edward; Schoot, Lotte; Kim, Minjae; Warnke, Lena; Kuperberg, Gina R. 2020. Neural evidence for the prediction of animacy features during language comprehension: Evidence from MEG and EEG Representational Similarity Analysis. *Journal of Neuroscience* 40.16: 3278-3291.
- Yeom, Seul-Ki; Fazli, Siamac; Müller, Kaus-Robert; Lee, Swong-Whan. 2014. An efficient ERP-based brain-computer interface using random set presentation and face familiarity. *PLOS ONE* 9.11: e111157.
-

¹Un galvanómetro es un instrumento utilizado para la medición de la corriente eléctrica.

Lo que esconden tus ojos: la metodología *eye-tracking* aplicada al estudio del lenguaje

Esther Álvarez García
Universidad de León
esther.alvarez@unileon.es

Resumen

El presente artículo ofrece una introducción a la metodología *eye-tracking* (o registro de movimientos oculares) aplicada al estudio del procesamiento del lenguaje. Concretamente, revisaremos cómo ciertos procesos cognitivos involucrados en el procesamiento lingüístico (principalmente, la activación léxica y el procesamiento oracional) se pueden reflejar (y, por tanto, estudiar) a través de los patrones de movimientos oculares. Aunque originalmente este método se empleó en tareas de lectura, en este trabajo analizaremos también su aplicación para el estudio del procesamiento del lenguaje desde otras órbitas, como la comprensión y la producción de mensajes orales.

Palabras clave: *eye-tracking*, procesamiento lingüístico, lectura, comprensión oral, producción oral.

Abstract

The present paper shows an introduction to the eye-tracking methodology applied to the study of language processing. Specifically, we review how certain cognitive processes involved in language processing (mainly, lexical activation and sentence processing) can be reflected (and, hence, studied) through eye-movement patterns. Even though this method was originally used in reading tasks, in this paper we also analyse its application to the study of language processing in other domains, such as the comprehension and production of oral messages.

Keywords: eye-tracking, language processing, reading, oral comprehension, oral production.

1. Introducción

Desde la Antigua Grecia existe la concepción de que los ojos son el espejo del alma, de que a través de una simple mirada somos capaces de revelar los más oscuros secretos de nuestro pensamiento. Lo cierto es que, en la actualidad, no existen evidencias de que esto sea así, pero de lo que sí existen pruebas –y, además, profusas– es de que los ojos son el espejo del lenguaje. El objetivo del presente trabajo será, por tanto, ilustrar la estrecha relación que mantienen los movimientos de los ojos con el lenguaje gracias al desarrollo de una técnica conocida como *eye-tracking* o, en español, “registro de movimientos oculares”. Desde hace varias décadas esta metodología ha permitido explorar algunos de los procesos cognitivos que nos permiten comprender y producir

mensajes lingüísticos a partir del análisis de los movimientos que los seres humanos realizamos con los ojos mientras leemos un texto u observamos una imagen.

Con este objetivo en mente, intentaremos responder algunas preguntas acerca de los fundamentos y aplicaciones de la metodología *eye-tracking*. En primer lugar, describiremos las premisas en las que se apoya esta metodología y su relación con el funcionamiento del sistema visual humano. Tras ello, comentaremos diferentes aspectos de carácter más técnico sobre los equipos que actualmente se emplean para realizar investigaciones con este método. Por último, revisaremos las principales contribuciones que la metodología *eye-tracking* ha realizado al estudio del lenguaje y cómo, a partir de ella, podemos relacionar ciertos movimientos oculares con ciertos procesos cognitivos.

2. ¿Qué es la metodología *eye-tracking*?

El *eye-tracking* o registro de movimientos oculares es una técnica experimental que consiste en seguir o rastrear la posición que adoptan los ojos mientras una persona realiza una tarea concreta, como leer un texto o buscar un objeto dentro de una imagen. Independientemente de la tarea, la premisa que subyace a esta metodología postula que los seres humanos dirigimos nuestra mirada hacia aquello que capta nuestra atención, de tal forma que aquello que miramos es, a su vez, lo que percibe y procesa el cerebro (Rayner 1998; Rayner y Liversedge 2011). Esta hipótesis, conocida como *ojo-mente* (Just y Carpenter 1980), se fundamenta en la propia anatomía del ojo humano. Los ojos son órganos que perciben la luz del entorno y, gracias al papel desempeñado por sus diferentes partes (pupila, córnea, cristalino, etc.), esa luz se proyecta sobre la retina creando una imagen invertida. En el centro de la retina se sitúa la fovea, un pequeño punto en torno al cual se concentra una gran cantidad de fotorreceptores. La fovea constituye, por tanto, el centro del campo visual, puesto que a través de ella percibimos una imagen precisa y detallada del estímulo que estemos observando. No obstante, la fovea tiene un tamaño reducido –se calcula que mide 1.5 milímetros de diámetro (Kolb et al. 1995)–, por lo que el campo visual que se percibe a través de ella es también limitado –se estima que la fovea abarca un campo visual de 1.2 grados (Hendrickson 2009)–. Así pues, el estímulo que se sitúe bajo la fovea se percibirá correctamente, mientras que todo aquello que quede fuera de este punto no se visualizará o se visualizará de manera menos precisa. Dada esta característica de la anatomía ocular, los seres humanos necesitamos mover los ojos con el objetivo de colocar la fovea sobre los diferentes estímulos que nos rodean si queremos percibirlos adecuadamente. A través de la fovea se proyectará una imagen nítida y precisa de esos estímulos y esta se transmitirá posteriormente al sistema nervioso para que el cerebro procese la información visual que han percibido nuestros ojos. Partiendo de esta hipótesis, el estudio de los patrones de movimientos oculares puede ayudarnos a descifrar algunos de los procesos cognitivos que subyacen a la interpretación y comprensión de la información visual.

Dentro de los movimientos oculares, podemos diferenciar dos patrones fundamentales. Por un lado, el punto sobre el que se coloca la fovea se conoce como *fijación*; por otro, el propio movimiento o salto que realizan los ojos para situar la fovea sobre un nuevo estímulo recibe el nombre de *sacada* o *movimiento sacádico*. Las fijaciones y las sacadas son las principales medidas de movimientos oculares¹ y se diferencian, entre

otros aspectos, por su función (Rayner 1998; Rayner y Pollatsek 2007; Rayner 2009; Raney et al. 2014): a través de una fijación extraemos información del estímulo que estamos observando; durante las sacadas, sin embargo, no es posible percibir información, ya que el movimiento que realiza el ojo ocurre a una gran velocidad –este fenómeno se conoce como *supresión sacádica* (Matin 1974; Wolverton y Zola 1983)–. Pese a ello, las sacadas cumplen una función fundamental: sitúan la fovea en una nueva región para así continuar percibiendo y procesando la información visual.

Aparte de su función, las fijaciones y las sacadas también difieren en sus unidades de medida: las fijaciones se miden en milisegundos, mientras que las sacadas se miden en grados –o, en el caso de la lectura, en número de caracteres–. La duración de una fijación, así como el tamaño de una sacada, depende del tipo de tarea que estemos ejecutando. Por ejemplo, durante la lectura de un texto, la duración de las fijaciones y el tamaño de las sacadas suelen ser menores respecto de una tarea en la que el participante deba visualizar una imagen o buscar un objeto dentro de ella. La Tabla 1 recoge la duración y la longitud medias de las fijaciones y las sacadas, respectivamente, en diferentes tareas desempeñadas por un hablante competente de inglés. No obstante, y como veremos en los próximos apartados, estos valores pueden modificarse significativamente en función del coste de procesamiento asociado a un estímulo: cuanto más costoso resulte procesar un estímulo, más largas serán las fijaciones y más cortas las sacadas.

Tarea	Duración media de la fijación (milisegundo)	Longitud media de la sacada (grado)
Lectura en silencio	225-250	2 (7-9 caracteres)
Lectura en voz alta	275-325	1.5 (6-7 caracteres)
Percepción de imágenes	260-330	4-5
Búsqueda visual	180-275	3

Tabla 1. Promedio de la duración de las fijaciones y de la longitud de las sacadas en diferentes tareas. Datos tomados de Rayner (2009); traducción mía

Un último aspecto que debemos tener en cuenta acerca del funcionamiento de los ojos es la composición del campo visual o perceptivo. Como señalamos previamente, el punto donde se sitúa la fovea durante una fijación constituye el centro del campo perceptivo, pero dentro de este podemos diferenciar tres regiones (Rayner 1998; Rayner y Pollatsek 2007; Rayner 2009). En primer lugar, el punto donde se coloca la fovea constituye la *región foveal* y es en ella donde se obtiene la visión más nítida, precisa y detallada de un estímulo. En el caso concreto de la escritura, la fovea se suele situar en torno a la primera mitad de una palabra, pues esta parece ser la localización óptima para poder reconocerla durante una única fijación. Esa posición óptima se conoce como *preferred viewing location* (Rayner 1979) y, de hecho, cuando la fovea no se sitúa en este punto durante la primera fijación, suele aumentar su duración (Inhoff et al. 2003; White et al. 2005a), así como la probabilidad de que la palabra reciba una nueva fijación (O'Regan y Lévy-Schoen 1987; McConkie et al. 1989).

En torno a la región foveal y en una posición más externa se sitúa la *región parafoveal*, en la que la percepción es menos nítida, pero aún es posible extraer cierta información visual. Concretamente, el procesamiento de los estímulos que se sitúan en la región parafoveal se conoce como *procesamiento parafoveal* y adquiere una relevancia

especial durante la lectura, por lo que se desarrollará con mayor detalle en los próximos subapartados.

La última región del campo perceptivo se conoce como *región periférica* y abarca todos aquellos estímulos que se sitúan más allá de la región parafoveal. En esta posición no es posible percibir información visual; por ejemplo, en el caso de la lectura no se pueden reconocer las palabras situadas en la región periférica. Así pues, la precisión visual mantiene una relación inversa con la posición de la fovea: a medida que un estímulo se aleja de la fovea, desciende la precisión visual con la que se percibe.

Un rasgo interesante de la región foveal –y, en general, del campo perceptivo– es que muestra una clara asimetría entre la cantidad de información que los ojos pueden percibir hacia la izquierda y hacia la derecha de una fijación. Como se puede observar en la Figura 1, en aquellas lenguas en las que se lee de izquierda a derecha, se calcula que la región foveal abarca 3-4 caracteres hacia la izquierda de una fijación (McConkie y Rayner 1976; Rayner et al. 1980; Underwood y McConkie 1985) y 7-8 caracteres hacia su derecha (Rayner y Bertera 1979; Rayner et al. 1982; Underwood y McConkie 1985). Esta asimetría se registra, igualmente, en la región parafoveal, que abarca, de nuevo, 3-4 caracteres hacia la izquierda de la fijación y unos 14-15 caracteres hacia su derecha (McConkie y Rayner 1975; Rayner y Bertera 1979; Rayner et al. 1982).

Los estudiantes colocaron una pancarta en la facultad.

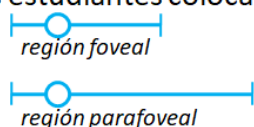


Figura 1. Asimetría perceptiva en las regiones foveal y parafoveal. El punto azul representa una fijación y las líneas azules, la extensión de las regiones foveal y parafoveal hacia el lado izquierdo y el derecho de esa fijación

Asimismo, llama la atención el hecho de que la asimetría del campo perceptivo se invierta en determinadas lenguas. Por ejemplo, Pollatsek et al. (1981) mostraron que en hebreo, una lengua en la que se lee de derecha a izquierda, el número de caracteres que se percibe hacia la izquierda de una fijación es mayor que hacia su derecha. Esta asimetría se registra también en la lectura vertical. Osaka y Oda (1991) advirtieron que, cuando un lector lee de arriba abajo en japonés, se percibe un mayor número de caracteres por debajo de una fijación que por encima de ella.

No obstante, esta desigualdad dentro del campo perceptivo puede verse reducida por diversos motivos. Por ejemplo, en aquellas lenguas con un sistema de escritura logográfico, como el japonés o el chino, la asimetría entre los dos lados de una fijación se reduce, de tal forma que, cuando se lee de izquierda a derecha, la región foveal abarca tan solo un carácter hacia la izquierda y 3-6 caracteres hacia la derecha (Ikeda y Saida 1978; Osaka 1987; Inhoff y Liu 1998). De manera similar, esta asimetría varía en función de la experiencia lectora: en los lectores competentes la región parafoveal puede alcanzar 14-15 caracteres hacia la derecha de una fijación; sin embargo, esta se ve reducida a un máximo de 10-12 caracteres en los lectores noveles (Rayner 1986). Por último, la diferencia entre los lados izquierdo y derecho del campo perceptivo también puede mitigarse como consecuencia de la complejidad del estímulo lingüístico o visual (Rayner 1986; Inhoff et al. 1989). Por ejemplo, cuando leemos una palabra desconocida

o poco frecuente en nuestra lengua, se reduce el número de caracteres percibidos hacia el lado derecho de una fijación.

3. ¿Cómo se registran los movimientos oculares?

El interés por analizar los movimientos oculares no es nuevo, sino que comenzó ya en el siglo XIX con los estudios de autores como Javal (1879), Lamare (1892) o Erdmann y Dodge (1898). Estos autores precisaron de un gran ingenio para desarrollar técnicas que permitiesen medir los movimientos de los ojos de una forma más rigurosa que a partir de la mera observación. Por ejemplo, Lamare ideó una técnica en la que un pequeño tubo se conectaba, por un lado, al párpado superior de los ojos y, por otro, a un tambor; de esta forma, podía escuchar las sacadas durante la lectura de un texto. De manera similar, Javal colocó pequeños espejos en los extremos de los ojos para registrar sus movimientos a través del reflejo. Pese a la indudable agudeza de estos autores y de las técnicas que concibieron, no será hasta los años 70 del siglo pasado cuando la metodología de movimientos oculares comience a alcanzar una mayor relevancia gracias al desarrollo de aparatos que permiten un registro más sencillo y eficaz (Rayner 1998; Wade y Tatler 2005).

Concretamente, el aparato que se emplea para rastrear los movimientos de los ojos se conoce con el nombre de *eye-tracker* –de momento no existe una adaptación de este término al español– y, aunque existen diferentes tipos, los más utilizados actualmente son los *video-based eye-trackers* (Wade y Tatler 2005; Raney et al. 2014; Carter y Luke 2020). Estos aparatos están compuestos por una cámara, que se encarga de grabar los movimientos de los ojos, y un iluminador de infrarrojos, que proyecta un rayo de luz infrarroja sobre los ojos del participante. Ese rayo de luz infrarroja se refleja, específicamente, sobre la pupila, creando una especie de punto luminoso, cuya posición será lo que registre la cámara del *eye-tracker*. Asimismo, el rayo de luz infrarroja crea un segundo reflejo sobre la córnea del ojo, que permanece estático y solo se desplaza en caso de mover la cabeza. De esta forma, el *eye-tracker* puede diferenciar los movimientos de los ojos que son consecuencia de mover la cabeza de aquellos que se producen con el objetivo de colocar la fovea en una nueva posición y que serán los que, verdaderamente, tengan interés científico.

Una característica fundamental de los *video-based eye-trackers* es su tasa de muestreo (Raney et al. 2014; Carter y Luke 2020). Este concepto hace referencia al número de veces por segundo que el aparato registra la posición del ojo y se mide en hercios (Hz). Así pues, si un *eye-tracker* tiene una tasa de muestreo de 1000 Hz, la posición del ojo se calculará mil veces por segundo –o, en otras palabras, una vez por milisegundo–. La importancia de esta característica reside en que la tasa de muestreo determina la precisión con la que se registran los movimientos oculares: cuanto mayor sea la tasa de muestreo, más preciso y fiable será el registro que se realice de esos movimientos. En la actualidad el *eye-tracker* con una mejor tasa de muestreo alcanza los 2000 Hz en visión monocular y los 1000 Hz en visión binocular. Dado que se asume que los dos ojos se mueven al mismo tiempo y en la misma dirección, los estudios de *eye-tracking* suelen rastrear los movimientos de un solo ojo –generalmente, del ojo dominante– para así evitar la pérdida de precisión (Hooge et al. 2018)².

Un segundo aspecto que permite diferenciar unos *eye-trackers* de otros es su disposición. Algunos aparatos son estáticos, de tal forma que no se pueden mover del laboratorio en el que están instalados. Este tipo de *eye-tracker* conlleva, generalmente, el uso de una mentonera sobre la que los participantes deben apoyar la cabeza mientras realizan el experimento para así evitar cualquier movimiento que pueda perjudicar o contaminar el registro. Por su parte, otros *eye-trackers* son portátiles, de tal manera que se pueden usar tanto dentro como fuera del laboratorio y, además, no requieren el uso de mentonera, lo que confiere una mayor libertad a los participantes –y, por tanto, una mayor naturalidad en la tarea–. El inconveniente de los *eye-trackers* portátiles se halla en su tasa de muestreo, pues se ve reducida considerablemente en comparación con la de los estáticos (Raney et al. 2014; Carter y Luke 2020).

Teniendo en cuenta todos estos aspectos, la pregunta que el lector podría plantearse es cómo escoger el *eye-tracker* idóneo para un estudio. Lo cierto es que la respuesta a este interrogante resulta bastante sencilla: aquel del que (por suerte) disponga el laboratorio o grupo de investigación en el que se trabaje. No obstante, en el caso de barajar diversas opciones, seleccionar uno u otro aparato dependerá de la tarea que se realice, de las características de la población con la que se trabaje o del tipo de información que se quiera obtener. Por ejemplo, en los estudios con bebés o niños resulta complejo, si no imposible, el uso de una mentonera, por lo que con este tipo de participantes se suele emplear los *eye-trackers* portátiles. De manera similar, si solo interesa obtener información espacial (es decir, saber dónde se mira), se puede emplear un *eye-tracker* con una menor tasa de muestreo (sea portátil o estático); en cambio, si se quiere obtener información temporal precisa (esto es, durante cuánto tiempo se mira), será necesario emplear un aparato con la mejor tasa de muestreo posible (generalmente, un *eye-tracker* estático).

4. ¿Qué nos enseñan los movimientos oculares acerca del lenguaje?

Desde hace varias décadas se ha empleado el registro de movimientos oculares en diferentes disciplinas como la sociología, el *marketing* o la medicina. No obstante, en el presente trabajo nos centraremos únicamente en analizar las principales aportaciones de esta metodología al campo de la psicolingüística o, en otras palabras, al estudio del procesamiento del lenguaje. A este respecto, la aplicación más lógica que parece tener la metodología *eye-tracking* consistiría en analizar los movimientos oculares durante una tarea de lectura, es decir, estudiar el lenguaje desde la órbita de la comprensión escrita. Como veremos, esta aplicación ha sido (y es) muy fructífera en el campo de la psicolingüística; sin embargo, no es la única, pues en los últimos años se ha desarrollado un nuevo paradigma conocido como paradigma del mundo visual (o *visual-world paradigm*). Este nuevo paradigma ha permitido aplicar la metodología del registro de movimientos oculares en tareas de comprensión y producción orales, lo que ha posibilitado estudiar el procesamiento del lenguaje desde estas otras perspectivas. A continuación, revisaremos las principales características y aportaciones de los estudios de *eye-tracking* desde la órbita de la lectura, en primer lugar, y de la comprensión y producción orales, en segundo lugar.

4.1. Movimientos oculares durante la lectura

La aplicación de la metodología *eye-tracking* en tareas de lectura se remonta a los albores de la propia técnica, pues autores como Javal (1879) o Lamare (1892) ya basaban sus estudios sobre los movimientos oculares en la lectura de frases o textos. No obstante, no fue hasta la década de los años 70 del siglo XX cuando este tipo de estudios adquirió una mayor relevancia gracias al desarrollo de equipos que permiten un registro más fácil y sofisticado de los movimientos oculares. Un segundo aspecto que facilitó la difusión de esta metodología en la investigación psicolingüística fue el hecho de que, frente a buena parte de los métodos conductuales, el registro de movimientos oculares permite estudiar el procesamiento del lenguaje de manera *online*, esto es, en el mismo momento en el que este ocurre. Por último, la forma de visualizar los estímulos en los experimentos de *eye-tracking* también supuso una gran ventaja respecto de paradigmas conductuales, ya que, con esta metodología, las frases o textos no aparecen fragmentados sino en su totalidad, lo que evita que se interrumpa el ritmo normal de lectura. Además, esta presentación de los estímulos permite que los lectores puedan volver hacia atrás y releer una palabra, un sintagma o toda una frase, un comportamiento habitual durante la lectura.

4.1.1. Medidas y unidades de análisis

En los estudios de movimientos oculares durante la lectura se manejan como unidades de análisis las fijaciones, por un lado, y las sacadas, por otro. En la Tabla 1 se recoge el promedio de la duración de las fijaciones y de la longitud de las sacadas en tareas de lectura tanto silenciosa como en voz alta. Como mencionamos anteriormente, estos valores medios pueden verse modificados en función del coste cognitivo que conlleva procesar un estímulo y que dependerá tanto de una serie de variables lingüísticas, que analizaremos a lo largo de este apartado, como de factores tipográficos. Por ejemplo, el uso de fuentes complejas o de un tamaño de letra pequeño suele provocar un aumento de las fijaciones y una reducción de la longitud de las sacadas (Rayner y Pollatsek 1989; Rayner et al. 2006).

Respecto a estas dos unidades de análisis, es importante considerar varios aspectos. Por una parte, durante la lectura, no todas las sacadas se realizan hacia regiones nuevas de una frase o de un texto, sino que es habitual volver hacia atrás y releer ciertas palabras. Esas sacadas que se dirigen hacia regiones de la frase o del texto que ya se habían leído reciben el nombre de *regresiones* y se calcula que en un lector competente el 10-15 % de las sacadas son regresiones (Rayner 1998; Rayner y Pollatsek 2007; Rayner 2009). Relacionado con este concepto, se encuentra un segundo: *refijación*. Así pues, cada vez que un lector realiza una regresión y posa su mirada sobre una palabra que había sido fijada previamente, hablaremos de *refijación*. Se considera que existen diferentes motivos por los que se puede producir una regresión y, en consecuencia, una refijación (Altmann et al. 1992; Rayner 1998; Rayner 2009). Por un lado, las regresiones –sobre todo si son cortas– pueden ser consecuencia de un error oculomotor: el sistema motor no planifica correctamente el movimiento de los ojos, de tal forma que estos se posan sobre una palabra diferente a la deseada, desencadenando una inmediata regresión para situar la fovea en la posición adecuada (Vitu et al. 1998). La segunda causa de una regresión es la complejidad lingüística de la frase o del texto: así, un lector suele regresar a regiones previas del texto o de la frase cuando no es capaz de integrar una

palabra dentro de la estructura oracional (Johnson et al. 2011) o cuando se percata de que la primera interpretación que había realizado de una estructura ambigua no es correcta (Frazier y Rayner 1982; Ferreira y Clifton 1986; Spivey y Tanenhaus 1998).



Figura 2. Movimientos oculares durante la lectura de una oración. Los puntos azules representan las fijaciones con su correspondiente distribución marcada a través de la numeración. La línea roja muestra una regresión y el punto rojo, una refijación

En cuanto a las fijaciones, es importante destacar que no todas son iguales y tampoco reflejan el mismo tipo de procesamiento lingüístico. Es por ello por lo que en los experimentos de lectura se suele diferenciar una serie de medidas de movimientos oculares, clasificadas en tempranas o tardías en función de si muestran un primer contacto de los ojos con el texto o, por el contrario, algún tipo de relectura³ (Rayner 1998; Clifton et al. 2007; Rayner y Liversedge 2011).

Medida	Definición	Tipo
Primera fijación (<i>first fixation</i>)	Duración de la primera fijación que recibe una palabra o región.	Temprana
Fijación única (<i>single fixation</i>)	Duración de la fijación cuando se produce una sola en una palabra o región.	Temprana
Primer barrido (<i>gaze duration</i> o <i>first-pass reading time</i> ⁴)	Suma de todas las fijaciones en una palabra o región desde que se entra en ella por vez primera hasta que se sale, ya sea por la derecha o la izquierda, también por vez primera.	Temprana
Probabilidad de primera fijación	Probabilidad de recibir una fijación durante el primer barrido o de rastrear inicialmente una palabra antes de una refijación.	Temprana
Cuasi primer barrido (<i>quasi first-pass</i> o <i>right-bounded reading time</i>)	Suma de todas las fijaciones en una palabra o región desde que se entra en ella por vez primera hasta que se sale por la derecha por vez primera. Por tanto, si después de una primera fijación se realiza una regresión hacia una región anterior y, tras ello, se produce una refijación en la palabra en cuestión, la duración de esta refijación se integrará en esta medida.	Temprana/tardía ⁵
Tiempo de lectura con regresiones (<i>go-past</i> o <i>regression path duration</i>)	Suma de todas las fijaciones en una palabra o región, así como de las refijaciones en palabras previas, desde que se entra en ella por vez primera hasta que se sale por la derecha por vez primera. A diferencia de la medida anterior, ahora se suman no solo las refijaciones en la palabra o región de interés, sino también en las anteriores a esta.	Temprana/tardía
Probabilidad de regresión	Probabilidad de realizar una regresión a una región anterior desde una determinada palabra.	Temprana/tardía
Segundo barrido (<i>second-pass reading time</i>)	Suma de la duración de todas las refijaciones en una palabra o región tras el primer barrido. En este caso, el lector ya ha salido de la región de interés por la derecha, pero regresa a ella para realizar una refijación.	Tardía
Tiempo total	Suma de la duración de todas las fijaciones en una palabra o región, incluyendo tanto el primer como el segundo barrido.	Tardía

Tabla 2. Medidas de movimientos oculares en tareas de lectura

Las medidas de movimientos oculares recogidas en la Tabla 2 han posibilitado el estudio, entre otros, de dos procesos diferentes involucrados en la comprensión de la lengua escrita: la activación léxica de las palabras y su integración en la estructura

sintáctico-semántica de una oración. En los siguientes subapartados, examinaremos qué medidas reflejan cada uno de estos dos procesos, así como las variables lingüísticas que las condicionan.

4.1.2. Activación léxica

Uno de los primeros pasos para poder comprender el significado de una frase o texto escrito es reconocer las palabras que en él aparecen. A través de los patrones de movimientos oculares, podemos estudiar cómo se activan las palabras en el lexicón mental y, sobre todo, qué variables lingüísticas condicionan su reconocimiento. Para ello, se parte, fundamentalmente, de las medidas tempranas de movimientos oculares.

Una de las principales variables que afecta al reconocimiento léxico de las palabras escritas es su longitud, que mantiene una relación proporcional con la duración de las medidas tempranas de movimientos oculares: así, cuanto más larga es una palabra, mayor es la duración de las fijaciones (Just y Carpenter 1980; Calvo y Meseguer 2002). De hecho, las palabras largas no suelen abarcarse en su totalidad durante una única fijación, por lo que, a medida que se incrementa su longitud, aumenta la posibilidad de que reciba una segunda fijación. Dada la influencia de esta variable en la duración de las fijaciones, se recomienda que aquellas palabras o regiones que se vayan a contrastar entre condiciones experimentales presenten la misma longitud⁶.

En segundo lugar, cuando se controla la longitud de las palabras, la frecuencia léxica se convierte en un factor determinante en su activación. Concretamente, esta variable mantiene una relación inversa con la duración de las medidas tempranas: cuanto más frecuente es una palabra, más fácil es su activación léxica y, en consecuencia, menor es la duración de las fijaciones (Just y Carpenter 1980; Rayner y Duffy 1986; Inhoff y Rayner 1986; White et al. 2018). No obstante, el efecto que produce la frecuencia léxica sobre los movimientos oculares se atenúa a medida que se repite una misma palabra. Así pues, la primera vez que se lee una palabra de baja frecuencia léxica se registra un aumento significativo de la duración de las fijaciones; sin embargo, si esa misma palabra aparece en más ocasiones a lo largo de la frase o del texto, la duración de sus fijaciones se reducirá e, incluso, puede llegar a no presentar diferencias respecto de una palabra de alta frecuencia léxica (Rayner et al. 1995).

Relacionada con la variable anterior se halla la predictibilidad, ya que aquellas palabras frecuentes en un determinado contexto se vuelven predecibles. Es por ello por lo que las palabras predecibles dentro de un contexto discursivo son más fáciles de activar y, en consecuencia, reciben fijaciones más cortas que aquellas que sean inesperadas (Ehrlich y Rayner 1981; Inhoff 1984; Balota et al. 1985).

La edad de adquisición es también una variable influyente en el reconocimiento léxico, pues aquellas palabras que el lector haya adquirido a una edad temprana reciben fijaciones más cortas que aquellas aprendidas más tardíamente (Juhasz 2005). De hecho, esta variable parece tener una mayor influencia en las medidas tempranas de movimientos oculares que la propia frecuencia léxica (Juhasz y Rayner 2003, 2006).

Por otra parte, la ambigüedad es una variable que también media en la duración de las fijaciones que recibe una palabra. En el caso de las palabras semánticamente ambiguas – es decir, aquellas que tienen dos o más significados –, sus valores pueden tener la misma frecuencia de aparición (*palabras equilibradas*) o diferente, es decir, uno de los

significados es más frecuente que el otro (*palabras sesgadas*). Duffy et al. (1988) mostraron que las palabras sesgadas reciben fijaciones más largas cuando el contexto previo favorece el significado menos frecuente, lo que sugiere que el significado más frecuente también se activa y compite con aquel por ser seleccionado. Por ejemplo, en español la palabra *banco* tiene diversos significados, algunos de ellos más frecuentes ('entidad financiera' o 'asiento') y otros menos habituales ('tierra situada junto a un río'). Si esta palabra apareciese en una frase como (1), las fijaciones sobre *bancos* serían más largas de lo habitual, pues, pese a que el contexto previo favorece el significado menos frecuente, aquellos valores más habituales también se activarían dificultando su reconocimiento léxico.

- (1) El ayuntamiento se encargó de limpiar el agua del río, pero los bancos seguían estando bastante sucios

Por su parte, las palabras equilibradas reciben fijaciones más largas cuando el contexto previo es neutro, indicativo ello de que, de nuevo, los significados compiten entre sí por recibir activación. Por ejemplo, en una oración como (2) la palabra *bancos* volvería a ser fijada durante un mayor tiempo, ya que el contexto previo a esta palabra no favorece ningún significado, de tal forma que los dos valores más frecuentes ('entidad financiera' y 'asiento') se activarían y rivalizarían entre sí por ser finalmente seleccionados.

- (2) Carlos se dirigió al banco y se sentó a leer a la sombra de un árbol

No obstante, no todas las palabras son ambiguas semánticamente, sino que la ambigüedad puede hallarse en otros niveles lingüísticos. Por ejemplo, Folk y Morris (2003) analizaron palabras cuya ambigüedad residía en su categoría gramatical, ya que podían funcionar como sustantivo –por ejemplo, *la copa de vino*– o como verbo –*mi primo vino*–, siendo una de estas dos interpretaciones más frecuente que la otra. En este caso, el hecho de que el contexto previo favoreciese la categoría menos frecuente no conllevaba un aumento de la duración de las fijaciones en medidas tempranas, lo que indica que esa interpretación menos frecuente era la única que se activaba pese a no ser la predilecta.

Asimismo, en lenguas con una ortografía opaca –es decir, aquellas en las que no hay una relación biunívoca entre los grafemas y los fonemas– la ambigüedad puede situarse en el nivel fonológico, por lo que la forma escrita de una palabra puede corresponderse con diferentes pronunciaciones. Por ejemplo, la palabra inglesa *wind* se puede articular como /wɪnd/ ('viento') o como /waɪnd/ ('enrollar'). Carpenter y Daneman (1981) analizaron este tipo de ambigüedad y registraron un aumento de la duración de las fijaciones en palabras ambiguas fonológicamente versus palabras no ambiguas. Ello indica que, durante la lectura de palabras ambiguas como *wind*, los dos códigos fonológicos se activan y compiten por ser finalmente seleccionados.

Por último, la estructura morfológica de una palabra también parece condicionar los patrones de movimientos oculares durante su reconocimiento. Por un lado, ciertos estudios han contrastado palabras compuestas cuyo primer o segundo lexema presentaba una mayor o menor frecuencia y han observado un aumento de las primeras fijaciones cuando el primer lexema del compuesto era menos frecuente, así como un aumento del primer barrido cuando el segundo lexema era menos frecuente (Hyönä y Pollatsek 1998;

Pollatsek et al. 2000; Juhasz et al. 2003). De manera similar, Juhasz (2007) analizó si el nivel de transparencia semántica del compuesto –es decir, si el significado global del compuesto refleja el de los morfemas que lo conforman o si, por el contrario, adquiere un valor completamente distinto– condiciona la duración de las medidas tempranas de movimientos oculares. En este caso, no se registraron diferencias entre los dos tipos de compuestos (transparentes versus opacos), pero sí se obtuvo un aumento de las fijaciones en palabras compuestas versus simples. Estos resultados –tanto los relativos a la frecuencia de los morfemas como a su transparencia semántica– parecen indicar que las palabras compuestas se dividen en sus partes individuales durante el proceso de activación léxica, pues, en caso contrario, no se registrarían diferencias en la duración de sus fijaciones respecto de palabras simples con la misma longitud y frecuencia léxica.

4.1.3. Integración sintáctico-semántica

Pese a la importancia que tiene la correcta activación léxica de las palabras, lo cierto es que los seres humanos no solemos comunicarnos a través de palabras aisladas, sino que estas suelen aparecer en unidades superiores, como oraciones o textos. A este respecto, se asume que el procesamiento lingüístico es *incremental* (Tanenhaus et al. 1995): a medida que leemos cada nueva palabra, intentamos establecer una relación entre esta y las palabras anteriores para construir una estructura sintáctico-semántica que nos permita comprender el mensaje. La metodología *eye-tracking* también se ha utilizado para analizar este proceso, es decir, para examinar cómo se integran las palabras dentro de la estructura oracional. Para ello, las medidas de movimientos oculares que se suelen analizar son diferentes a las señaladas antes para la activación léxica, excepto una: el primer barrido. Aunque esta medida se relaciona con el reconocimiento léxico, el hecho de que una misma palabra reciba más de una fijación antes de salir de ella puede reflejar, igualmente, el coste de integrarla dentro de la estructura oracional (Rayner et al. 1989). Además del primer barrido, este proceso se puede examinar a partir de las medidas intermedias y tardías.

Los primeros estudios de movimientos oculares que se interesaron por analizar este fenómeno buscaban obtener evidencias empíricas que apoyasen los postulados de las teorías de corte modular sobre el procesamiento del lenguaje. Dentro de estas teorías, aquella que ha tenido una mayor repercusión en el campo de la psicolingüística es el modelo de vía muerta o *garden-path* (Frazier y Rayner 1982; Frazier 1987). Este modelo defiende un procesamiento encapsulado de la información lingüística y, en el caso concreto de las oraciones, aboga por un procesamiento en dos estadios independientes: en el primer estadio se construiría la estructura sintáctica de la oración, teniendo para ello en cuenta la categoría gramatical de las palabras y un conjunto de principios o reglas de naturaleza también sintáctica que se aplicarían de manera ciega –es decir, sin considerar otro tipo de información lingüística–; tras ello, se iniciaría el segundo estadio, en el que se activaría la información semántico-pragmática correspondiente. Para poner a prueba esta hipótesis, los estudios empíricos empleaban oraciones sintácticamente ambiguas a partir de las cuales se intentaba determinar si el cerebro –o, como se suele denominar en estos estudios, el procesador o *parser*– era capaz de detectar su ambigüedad de manera temprana. Así pues, si el cerebro solo tuviese en cuenta la información sintáctica durante el primer estadio, la ambigüedad estructural pasaría desapercibida y la interpretación que se activaría de esa oración sería

incorrecta, lo que se conoce como “caer en una vía muerta”; si, por el contrario, de manera temprana se activara otro tipo de información lingüística (diferente de la sintáctica), el cerebro podría detectar la ambigüedad de la oración durante el primer estadio y, por tanto, darle una correcta interpretación.

Uno de los primeros estudios de este tipo fue llevado a cabo por Frazier y Rayner (1982), quienes analizaron los movimientos oculares durante la lectura de oraciones ambiguas como (3):

- (3) The second wife will claim the inheritance belongs to her
'La segunda esposa reivindicará (que) la herencia le pertenece'

En la oración (3) la región ambigua –y, por tanto, de interés– es el sintagma *the inheritance*, ya que, *a priori*, podría funcionar como complemento directo del verbo principal o como sujeto de una oración subordinada, siendo esta última la interpretación correcta. Frazier y Rayner (1982) obtuvieron dos patrones de movimientos oculares complementarios. Por un lado, no registraron ninguna diferencia en los tiempos de lectura de la región de interés, lo que indicaría que el cerebro no detecta la ambigüedad de la oración a la altura de esta región y que, por tanto, selecciona la estructura más simple: aquella en la que la región ambigua funciona como complemento directo. Por su parte, en la región desambiguadora –en el ejemplo (3), *belongs*–, se produjo un incremento de la duración de las fijaciones de primer y segundo barrido, así como un aumento de las regresiones hacia regiones anteriores. Este segundo patrón confirmaría que, en un primer momento, los participantes seleccionan la interpretación incorrecta de la oración y que solamente cuando alcanzan la región desambiguadora, detectan el error, lo que conlleva un aumento de los tiempos de lectura, así como de las regresiones, que se correspondería con el proceso de reanálisis de la estructura oracional. En consecuencia, Frazier y Rayner interpretaron sus resultados a favor del modelo de vía muerta (Rayner et al. 1983; Ferreira y Clifton 1986; Ferreira y Henderson 1990).

Con posterioridad estos estudios empezaron a incluir nuevas variables con el objetivo de analizar si estas condicionaban el procesamiento de la estructura oracional y si ello se veía, a su vez, reflejado en ciertos patrones de movimientos oculares. A este respecto, las variables que influyen en el proceso de integración oracional no están definidas de manera tan clara como para el caso de la activación léxica, ya que los resultados obtenidos hasta la fecha no siempre concuerdan. Esta falta de sistematicidad parece deberse, entre otros motivos, a la disparidad del material empleado –se han utilizado diferentes estructuras–, a la diferencia de longitud de la región de interés –en algunos casos abarca una sola palabra mientras que, en otros, varias– o al tipo de información lingüística manipulada (Clifton et al. 2007). Es, concretamente, este último aspecto –la información lingüística– el que tendremos en cuenta a continuación para clasificar las posibles variables que influyen en los movimientos oculares durante la lectura y el procesamiento de oraciones y que agruparemos en tres tipos: la frecuencia, los rasgos semánticos y el contexto discursivo.

Una de las primeras variables analizadas ha sido la frecuencia. Por un lado, ciertos autores han examinado si la mayor o menor frecuencia léxica de una palabra puede condicionar, aparte de su activación, cómo se establece una relación sintáctica entre esta y otras palabras de la oración. Por ejemplo, Rayner et al. (1989) contrastaron oraciones

en las que un mismo sustantivo podía estar precedido por un adjetivo de mayor –*electric* en (4a)– o menor frecuencia léxica –*acoustic* en (4b)–.

- (4) a. Jerry’s electric guitar needs new strings
‘La guitarra eléctrica de Jerry necesita cuerdas nuevas’
b. Jerry’s acoustic guitar needs new strings
‘La guitarra acústica de Jerry necesita cuerdas nuevas’

Los resultados mostraron que los adjetivos de menor frecuencia recibían primeras fijaciones más largas que aquellos de mayor frecuencia, lo que reflejaría un mayor coste en su reconocimiento. Más interesante, sin embargo, es el hecho de que el sustantivo que aparecía tras dichos adjetivos –*guitar* en el ejemplo (4)– también recibía fijaciones más largas tras *acoustic* versus *electric* durante el primer barrido. Dado que el sustantivo es el mismo en ambas condiciones, esta diferencia no podía ser consecuencia de su activación léxica, sino que este efecto reflejaría un mayor coste a la hora de establecer una relación sintáctica entre el adjetivo de menor frecuencia y el sustantivo. Esta misma conclusión dedujeron Johnson et al. (2011), aunque en su estudio el mayor coste a la hora de integrar palabras de menor versus mayor frecuencia se tradujo en un aumento de las fijaciones en la medida de cuasi primer barrido. Así pues, la frecuencia léxica de una palabra parece condicionar su integración en la estructura oracional (White et al. 2018) y la existencia de cierta dificultad durante este proceso se puede reflejar a través de dos patrones de movimientos oculares: fijar durante un mayor tiempo la palabra contigua (Rayner et al. 1989) o regresar a regiones anteriores y refijar durante un mayor tiempo la palabra problemática (Johnson et al. 2011).

No obstante, en los estudios sobre procesamiento oracional no solo se puede manipular la frecuencia léxica de las palabras, sino también la frecuencia con la que estas se combinan con algunas estructuras sintácticas –es decir, la frecuencia combinatoria–. Por ejemplo, Ferreira y Henderson (1990) replicaron el estudio de Frazier y Rayner (1982), pero controlaron la frecuencia combinatoria de los verbos: algunos se combinaban más frecuentemente con un sustantivo en la función de complemento directo –por ejemplo, el verbo *saw* en (5a)– mientras que otros aparecían con mayor frecuencia combinados con una oración subordinada –*realized* en (5b)–.

- (5) a. John saw the road was closed
‘John vio (que) la carretera estaba cerrada’
b. John realized the road was closed
‘John advirtió (que) la carretera estaba cerrada’

Los resultados mostraron que los verbos que se combinan más frecuentemente con sustantivos (5a) recibían fijaciones más largas que aquellos que suelen aparecer con una oración subordinada (5b), pero esta diferencia solo se registraba en la medida tardía de tiempo total. Estudios posteriores han señalado, sin embargo, que la frecuencia combinatoria puede condicionar el procesamiento sintáctico no solo de manera tardía, sino también en medidas tempranas como el primer barrido (Trueswell et al. 1993; Garnsey et al. 1997; Wilson y Garnsey 2009).

Más allá de la frecuencia, numerosos estudios dentro del campo de la psicolingüística se han interesado en analizar mediante la metodología *eye-tracking* si los rasgos semánticos de los constituyentes de una oración condicionan la interpretación que se hace de esta. Estos estudios surgieron, en un principio, para poner a prueba las hipótesis de las teorías modulares comentadas previamente: se manipulaban los rasgos semánticos de algún constituyente para examinar si ello ayudaba a los lectores a no caer en una vía muerta o, en caso de caer, si aliviaba el coste cognitivo asociado con su reanálisis. Por ejemplo, Ferreira y Clifton (1986) examinaron oraciones como las del ejemplo (6), cuya ambigüedad reside en que el primer verbo *–examined–* puede ser analizado como el verbo de la oración principal o como el participio de pasado de una oración de relativo reducida⁷, siendo esta última la interpretación correcta. Asimismo, estos autores manipularon el rasgo semántico de la animacidad del antecedente, pudiendo ser animado (6a) o inanimado (6b).

- (6) a. The defendant examined by the lawyer turned out to be unreliable
'El acusado examinó/examinado por el abogado resultó ser poco fiable'
- b. The evidence examined by the lawyer turned out to be unreliable
'La evidencia examinó/examinada por el abogado resultó ser poco fiable'

Los resultados mostraron un aumento de las fijaciones durante el primer y segundo barrido en oraciones ambiguas como las de (6) versus oraciones no ambiguas. Más interesante es el hecho de que no se registraron diferencias en función del rasgo de la animacidad en ninguna de las medidas de movimientos oculares, lo que indicaría que esta información semántica no influye en el procesamiento oracional. No obstante, y como adelantamos previamente, los resultados en este tipo de estudios no siempre son sistemáticos: por ejemplo, Trueswell et al. (1994) replicaron el estudio de Ferreira y Clifton (1986) al considerar que los materiales empleados por estos estaban sesgados. En su trabajo Trueswell et al. sí obtuvieron diferencias significativas como consecuencia del rasgo semántico del antecedente: la medida de primer barrido presentaba fijaciones más largas en oraciones ambiguas en comparación con las no ambiguas cuando el antecedente era animado (6a); en cambio, con antecedentes inanimados (6b) no se registraban diferencias en la duración de las fijaciones entre las oraciones ambiguas y no ambiguas. Estos últimos resultados indicarían, por tanto, que la información semántica puede condicionar el procesamiento sintáctico de una oración y facilitar (o no) el reconocimiento de su estructura.

Aunque la mayoría de los trabajos que han analizado la influencia de los rasgos semánticos parte de oraciones sintácticamente ambiguas (Binder et al. 2001; Clifton et al. 2003; Hoeks et al. 2006), algunos de ellos también han considerado oraciones no ambiguas y, de hecho, es con este tipo de estructuras con las que se han obtenido resultados más sistemáticos. Muchos de estos estudios parten de la famosa asimetría registrada entre dos tipos de oraciones de relativo: así, se suele considerar que aquellas estructuras en las que el antecedente funciona como sujeto de la oración relativa (7a) son más fáciles de procesar que aquellas en las que dicho constituyente funciona como objeto (7b) (Holmes y O'Reagan 1981; Staub 2010).

- (7) a. The director that watched the movie received a prize
'El director que vio la película recibió un premio'

- b. The director that the movie pleased received a prize
'El director (al) que la película gustó recibió un premio'
- (8) a. The movie that pleased the director received a prize
'La película que gustó al director recibió un premio'
- b. The movie that the director watched received a prize
'La película que el director vio recibió un premio'

Lo interesante de estas oraciones de relativo es el hecho de que permiten manipular el rasgo semántico de la animacidad del antecedente sin modificar la estructura sintáctica y sin que esta sea ambigua, por lo que, de esta forma, se puede analizar la posible influencia de la información semántica en el procesamiento sintáctico de una manera más natural en comparación con los estudios previos que intentaban validar las teorías de corte modular. En este sentido, se ha demostrado que la asimetría entre las oraciones de relativo se puede neutralizar cuando el antecedente hace referencia a un ente inanimado (8). Así pues, una oración de relativo recibe fijaciones más largas en medidas intermedias y tardías cuando el antecedente animado funciona como objeto frente a la posición de sujeto; sin embargo, cuando el antecedente es inanimado, no se registran diferencias en la duración de las fijaciones debido a la función de sujeto u objeto del antecedente (Traxler et al. 2002; Mak et al. 2002, 2006; Traxler et al. 2005; Betancort et al. 2009). Estos resultados apoyarían los obtenidos en trabajos previos con estructuras ambiguas e indicarían que, al menos en las medidas intermedias y tardías de movimientos oculares, es posible registrar cierta influencia de la información semántica sobre el procesamiento de una estructura oracional.

Por último, la influencia de los rasgos semánticos también se ha analizado a partir de estructuras en las que se manipula el nivel de plausibilidad semántico-pragmática de alguno de sus constituyentes. Por ejemplo, Rayner et al. (2004) contrastaron oraciones en las que la palabra de interés –*carrots* en el ejemplo (9)– podía ser plausible semánticamente dentro de una oración (9a), poco probable pero plausible (9b) o completamente implausible (9c).

- (9) a. John used a knife to chop the large carrots for dinner
'Juan usó un chuchillo para cortar las zanahorias grandes de la cena'
- b. John used an axe to chop the large carrots for dinner
'Juan usó un hacha para cortar las zanahorias grandes de la cena'
- c. John used a pump to inflate the large carrots for dinner
'Juan usó una bomba para inflar las zanahorias grandes de la cena'

Los resultados mostraron que la condición anómala (9c) producía un aumento inmediato de las fijaciones durante el primer barrido mientras que, en la condición poco probable pero plausible (9b), el efecto se retrasaba hasta la medida intermedia de tiempo de lectura con regresiones. Estos resultados indican que los movimientos oculares son sensibles al nivel de plausibilidad semántico-pragmática de los constituyentes de una oración, de tal forma que cuanto menor sea este nivel, más fácil y rápidamente se detectará el error (Warren et al. 2008; Rayner y Liversedge 2011).

El tercer y último tipo de información lingüística que revisaremos en relación al procesamiento oracional es el contexto discursivo. Diversos estudios de movimientos

oculares han manipulado la información contextual con el objetivo de determinar si esta también influye en el proceso de crear e integrar las palabras en una estructura oracional. Concretamente, esta cuestión se ha examinado a partir de dos fenómenos lingüísticos: la resolución de ambigüedades sintácticas y el establecimiento de relaciones anafóricas.

Por un lado, diversos trabajos han examinado si el hecho de presentar cierta información contextual previa a una oración ambigua puede ayudar a evitar la ambigüedad y si ello se reflejaría en los patrones de movimientos oculares. Por ejemplo, Spivey y Tanenhaus (1998) contrastaron oraciones como (10c), cuya ambigüedad reside en el primer verbo –*selected*–, ya que puede ser interpretado como el verbo de la oración principal o como el participio de pasado de una oración de relativo reducida. Antes de esta oración ambigua, se presentaba un contexto que podía ser nulo, como el de (10a), o relevante, como el de (10b):

- (10) a. An actress and the producer's niece were auditioning for a play. The director selected the actress but not the niece
'Una actriz y la sobrina del productor participaban en el casting para una obra. El director seleccionó a la actriz, pero no a la sobrina'
- b. Two actresses were auditioning for a play. The director selected one of the actresses but not the other
'Dos actrices participaban en el casting para una obra. El director seleccionó a una de ellas, pero no a la otra'
- c. The actress selected by the director believed that her performance was perfect
'La actriz seleccionó/seleccionada por el director creía que su actuación había sido perfecta'

La diferencia entre un contexto y otro reside en que el primero de ellos (10a) carece de información discursiva que facilite la correcta interpretación de la oración ambigua mientras que el segundo (10b) sí posee dicha información: concretamente, el contexto de (10b) presenta dos referentes compatibles con el sintagma *the actress* en (10c), de tal forma que esta información contextual ayuda a interpretar el verbo ambiguo –*selected* en (10c)– como el participio de pasado de una oración relativa reducida, cuya función será, precisamente, la de identificar a cuál de los dos referentes presentados en el contexto previo se alude.

Los resultados de Spivey y Tanenhaus (1998) mostraron que, tras un contexto nulo como el de (10a), las oraciones ambiguas recibían fijaciones más largas que las oraciones no ambiguas y se producía, asimismo, un mayor número de regresiones desde la región desambiguadora –*by the director* en (10c)– hacia las anteriores. Sin embargo, cuando se presentaba un contexto con dos posibles referentes, como el de (10b), no se registraban diferencias significativas entre las oraciones ambiguas y las no ambiguas. Así pues, estos resultados revelan que la información contextual condiciona cómo se integran las palabras en una oración, facilitando su correcta interpretación en caso de ambigüedad (Altmann et al. 1992, 1994; Hoeks et al. 2002). Igualmente interesante es el hecho de que este efecto se evidencie ya en medidas tempranas de movimientos

oculares, como el primer barrido, lo que demostraría la pronta influencia de la información contextual en el procesamiento oracional.

Por su parte, varios estudios también han analizado la influencia de la información contextual examinando cómo se establece una relación anafórica entre el antecedente y la unidad que remite a este. En este caso, las variables que generalmente se han manipulado son, en primer lugar, la distancia entre las unidades correferenciales: así, el sustantivo que funciona como antecedente y el pronombre que remite a él pueden aparecer contiguos –*Susan* y *she* en (11a)– o separados por varios constituyentes de la oración –*Mark* y *he* en (11b)–.

- (11) a. A group of people who shared an interest in photography had recently started writing a newsletter of their activities. In fact, in one room Mark was mailing a copy of the paper to Susan. She was very involved in photography and spent every weekend taking pictures
'Un grupo de gente que compartía interés por la fotografía había empezado recientemente a escribir una circular sobre sus actividades. De hecho, en una habitación Mark estaba enviando una copia del documento a Susan. Ella estaba muy involucrada en la fotografía y se pasaba cada fin de semana sacando fotos'
- b. A group of people who shared an interest in photography had recently started writing a newsletter of their activities. In fact, in one room Mark was mailing a copy of the paper to Susan. He was very involved in photography and spent every weekend taking pictures
'Un grupo de gente que compartía interés por la fotografía había empezado recientemente a escribir una circular sobre sus actividades. De hecho, en una habitación Mark estaba enviando una copia del documento a Susan. Él estaba muy involucrado en la fotografía y se pasaba cada fin de semana sacando fotos'

La segunda variable que se ha tenido en cuenta en estos estudios es el grado de tipicidad del segundo elemento respecto del antecedente: por ejemplo, si el antecedente es el sustantivo *pájaro*, el segundo segmento puede ser una unidad más prototípica dentro de esta categoría semántica –*jilguero*– o menos –*vencejo*–. Aunque no todos los resultados concuerdan, existen evidencias de que la duración de las fijaciones –tanto de medidas tempranas como tardías– se incrementa cuanto mayor es la distancia entre el antecedente y la unidad correferente (Ehrlic y Rayner 1983). De manera similar, se registra un incremento de las fijaciones cuando la unidad correferente es menos prototípica dentro de una determinada categoría semántica (Duffy y Rayner 1990) o menos esperable en relación con nuestro conocimiento del mundo (Sturt 2003; Duffy y Keir 2004). Respecto a estas variables, es también interesante el hecho de que varios de estos estudios han obtenido un efecto de arrastre o *spillover*, esto es, la diferencia entre unas y otras condiciones no solo se registra en la región *target* –es decir, aquella en la que aparece la unidad correferente como, por ejemplo, el pronombre personal *she* o *he* en (11a) y (11b), respectivamente–, sino también en la región posterior a esta (Ehrlic y Rayner 1983; Duffy y Rayner 1990; Sturt 2003; Duffy y Keir 2004). Este efecto revela que, aunque el cerebro comience a establecer la relación anafórica entre dos palabras cuando los ojos se posan sobre la segunda de ellas, este proceso no se completaría de

manera inmediata, sino que continuaría mientras los participantes fijan las siguientes regiones.

En resumen, en los dos últimos apartados hemos intentado mostrar la relación existente entre las medidas de movimientos oculares tradicionalmente empleadas en los estudios de lectura y los procesos cognitivos que estas reflejan. No debemos olvidar, sin embargo, que esta relación no es perfecta ni unívoca –especialmente en el caso de la integración de las palabras en la estructura oracional–, por lo que los futuros trabajos deberán seguir indagando sobre cómo los diferentes tipos de información lingüística condicionan la comprensión del lenguaje escrito para así obtener una mejor interpretación de la relación ojo-mente (Rayner y Liversedge 2011).

4.1.4. Procesamiento parafoveal durante la lectura

Antes de cerrar el apartado dedicado a la aplicación de la metodología *eye-tracking* al estudio de la lectura, debemos realizar un par de puntualizaciones acerca del procesamiento parafoveal. Como señalamos previamente, la visión en la región parafoveal es menos nítida y precisa que en la región foveal, pero aun así es posible percibir cierta información lingüística acerca de las palabras que se localizan en ella. Examinar este aspecto es fundamental, pues si una palabra se procesa de manera parafoveal, se reducirá su tiempo de lectura cuando, finalmente, se fije en la fovea. Esa reducción de su tiempo de lectura puede acarrear la falsa idea de que una palabra es fácil de comprender, cuando, en realidad, dicha reducción podría deberse a un procesamiento parafoveal. Dicho de otro modo, una palabra puede conllevar un alto coste de procesamiento, pero si una parte de sus rasgos ha sido activada parafovealmente, entonces la duración de sus fijaciones se reducirá. A este respecto, diversos experimentos de registro de movimientos oculares⁸ han constatado que el procesamiento parafoveal de una palabra se circunscribe a ciertos rasgos lingüísticos: existen evidencias de que tanto la forma ortográfica (Rayner et al. 1982; Inhoff 1989a) como fonológica (Pollatsek et al. 1992; Henderson et al. 1995) de una palabra se pueden activar cuando esta todavía se sitúa en la región parafoveal; sin embargo, no parece ocurrir lo mismo con su frecuencia léxica (Carpenter y Just 1983; Henderson y Ferreira 1993), su estructura morfológica (Lima 1987; Inhoff 1989b) o sus rasgos semánticos (Rayner et al. 1986; Altarriba et al. 2001).

Más allá del tipo de información lingüística que se pueda procesar parafovealmente, ciertos trabajos han demostrado que la cantidad de información también está restringida y supeditada al coste de comprender la palabra que se esté fijando: cuanto más costosa sea la activación léxica o la integración oracional de la palabra situada en la región foveal, menor será el número de recursos cognitivos disponibles para iniciar el procesamiento parafoveal y, por tanto, menor será la cantidad de información que se pueda reconocer sobre la palabra localizada en esta segunda región (White et al. 2005b; Drieghe 2008).

Una última consecuencia del procesamiento parafoveal es que, en ciertas situaciones, las palabras que se sitúan en esta región pueden reconocerse completamente, lo que conlleva un salto desde la palabra fijada hasta la segunda por detrás de esta, de tal forma que la palabra situada entre ellas –es decir, la palabra situada en la región parafoveal– no recibe ninguna fijación. Generalmente, la probabilidad de saltar una palabra es mayor cuando esta es corta (Drieghe et al. 2004), de alta frecuencia léxica (Rayner et al.

1996) y predecible en un determinado contexto (Drieghe et al. 2004). Como es de suponer, estas características se combinan con mayor asiduidad en las palabras función que en las palabras de contenido, por lo que las primeras se suelen saltar más y fijar menos que las segundas: las palabras función suelen recibir un 35 % de fijaciones frente al 85 % de las palabras de contenido (Rayner 1998; Rayner y Pollatsek 2007; Rayner 2009). No obstante, el hecho de que una palabra no se fije no implica que no se procese (Fisher y Shebilske 1985); al contrario, esa palabra se habrá procesado de manera parafoveal. Como consecuencia, cuando una palabra se salta, se suele registrar un incremento de la duración de la fijación anterior (Pollatsek et al. 1986; Rayner et al. 2003).

4.2. Movimientos oculares durante la comprensión y producción orales

Dentro del campo de la psicolingüística la metodología *eye-tracking* también se puede emplear para estudiar los procesos cognitivos involucrados en la comprensión y producción del lenguaje oral. Concretamente, el estudio de estos procesos a través de los movimientos oculares se hizo posible gracias a la aparición del paradigma conocido como del mundo visual (o *visual-world paradigm*). El primero en emplear este paradigma para estudiar la comprensión de mensajes orales fue Cooper en 1974; sin embargo, su uso no se extendió en el ámbito de la psicolingüística hasta la década de los 90. En los próximos apartados analizaremos, por un lado, las principales características del paradigma del mundo visual y, por otro, los procesos cognitivos que podemos estudiar a partir de él tanto en el ámbito de la comprensión como de la producción lingüísticas.

4.2.1. Rasgos y unidades de análisis

El paradigma del mundo visual consiste en exponer a los participantes una escena en la pantalla del ordenador a la vez que realizan una tarea de comprensión o de producción lingüística y se registran sus movimientos oculares sobre la escena que están visualizando. En el caso de la comprensión, los participantes escucharán una palabra o una frase que se corresponde con aquel elemento de la escena que deben mirar; esta práctica se puede combinar con dos tareas: (i) los participantes escuchan el estímulo acústico mientras observan la escena (*look-and-listen task* o *passive listening*) o (ii) realizan una tarea siguiendo las instrucciones que se dan a través del estímulo acústico y que suele consistir en seleccionar una imagen o arrastrarla (*action-based task*) (Tanenhaus y Trueswell 2006; Salverda y Tanenhaus 2017). Por su parte, cuando este paradigma se aplica al estudio de la producción lingüística, los participantes deben articular una palabra o una frase que describa la escena que están visualizando.

Generalmente, la escena consta de varios objetos que se pueden organizar de dos formas diferentes: como imágenes independientes dentro de la pantalla o como elementos que simulan una situación semirrealista (Figura 3). Aquel objeto que los participantes deben mirar durante la comprensión o producción de una palabra o una frase es el estímulo *target* mientras que el resto son distractores. Algunos de estos distractores están relacionados lingüísticamente con el *target*, de tal forma que son sus competidores – fonológicos, semánticos, léxicos, etc.–; otros, en cambio, no guardan ninguna relación con el *target* (Salverda y Tanenhaus 2017). En cualquier caso, es importante que los objetos se coloquen a cierta distancia unos de otros para evitar el solapamiento de las

fijaciones y, asimismo, deben aleatorizarse entre los diferentes ensayos para impedir que los participantes conozcan de antemano la posición en la que aparecerá el elemento *target*.

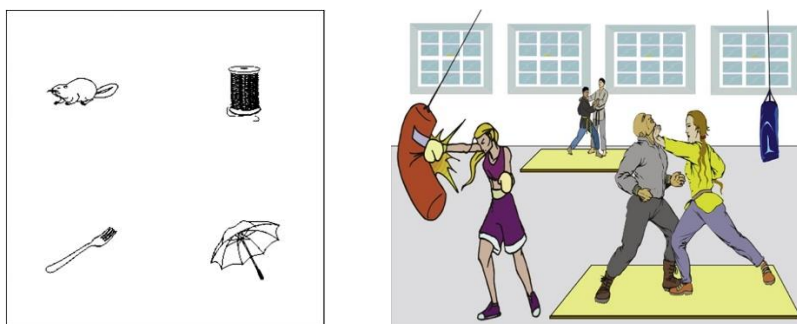


Figura 3. El estímulo visual de la izquierda muestra cuatro objetos independientes dentro de la escena – tomado de Huettig y McQueen (2007)–; el estímulo visual de la derecha representa una escena semirrealista –tomado de Humphreys et al. (2016)–

En cuanto al estímulo acústico, se recomienda incluir un intervalo de unos milisegundos entre la aparición de la escena en la pantalla y el inicio de la palabra o frase, ya que, de esta forma, los participantes disponen de un tiempo para familiarizarse con los objetos de la escena y no distraerse con ellos cuando comience el estímulo acústico. De hecho, algunos estudios apuntan que esa previsualización también es importante para obtener los efectos deseados, pues, en caso contrario, los participantes dirigen directamente su mirada hacia el elemento *target* (Dahan et al. 2007). Pese a haber una previsualización, los movimientos oculares que interesan son aquellos que se corresponden con el estímulo acústico, por lo que, una vez realizado el registro, se seleccionarán las ventanas temporales en las que se escuchó o articuló la palabra de interés y se analizarán los movimientos oculares realizados durante ese preciso momento, así como con posterioridad a él.

Partiendo de este escenario, las unidades de análisis que se suelen manejar en los experimentos que emplean el paradigma del mundo visual son, en cierto sentido, diferentes a las utilizadas en los estudios de lectura: (i) la proporción de fijaciones que recibe cada uno de los objetos de la escena, (ii) la duración de dichas fijaciones, (iii) la dirección y tiempo que conlleva realizar una sacada hacia el *target* o alguno de los distractores y (iv) la probabilidad de realizar una sacada hacia cualquiera de los objetos. A partir de estas medidas no se busca evaluar el coste cognitivo, como en los estudios de lectura, sino que se pretende inferir cómo es la representación mental que los participantes crean a medida que escuchan o producen una palabra o una frase (Tanenhaus 2007).

4.2.2. Comprensión lingüística

El paradigma del mundo visual se aplicó en un primer momento al estudio de la comprensión oral: a medida que los participantes escuchan un estímulo acústico, se registran sus movimientos oculares durante la inspección de una imagen. Gracias a esta aplicación se ha podido estudiar, entre otros, dos procesos fundamentales en la comprensión de los mensajes lingüísticos: la activación léxica y el procesamiento sintáctico.

Por un lado, los estudios sobre activación léxica se han interesado en analizar si, durante el reconocimiento de una palabra, median rasgos fonológicos, semánticos o ambos. Por ejemplo, Allopenna et al. (1998) emplearon el paradigma del mundo visual para contrastar objetos cuyos nombres podían tener una forma fonológica similar. Así, en la escena se incluía el elemento *target* (*beaker* ‘vaso’), un competidor fonológico (*beetle* ‘escarabajo’), un competidor con una entonación similar al *target* (*speaker* ‘altavoz’) y un distractor no relacionado (*carriage* ‘carricoche’). Mientras los participantes inspeccionaban la escena, escuchaban una frase en la que aparecía la palabra *target*: *Pick up the beaker...* Los resultados mostraron que, durante la escucha de la primera sílaba (/bi:/), los participantes repartían de manera equitativa sus fijaciones entre el elemento *target* y su competidor fonológico (*beetle*); sin embargo, en el momento en el que se percibía el fonema desambiguador (/k/) aumentaban las fijaciones hacia el *target* y disminuían las dirigidas hacia el competidor fonológico. Estos resultados sugieren que, durante el reconocimiento oral de una palabra, no solo se activa aquella intencionada, sino también las que presentan una forma fonológica similar (Swingley et al. 1999).

Huettig y McQueen (2007) obtuvieron resultados similares, aunque estos autores incluyeron, aparte de competidores fonológicos, competidores semánticos y visuales en su estudio. Así, el elemento *target* (*beaker*) contrastaba con un competidor fonológico (*beaver* ‘castor’), un competidor semántico que pertenecía a la misma categoría (*fork* ‘tenedor’) y un competidor visual cuya forma era similar (*bobbin* ‘bobina’). Los resultados mostraron que, durante los primeros 400 ms tras el inicio de la palabra *target*, los participantes dirigían su mirada tanto al *target* como al competidor fonológico (*beaver*); en cambio, en las siguientes ventanas temporales eran los competidores semántico (*fork*) y visual (*bobbin*) los que recibían una mayor proporción de fijaciones respecto del competidor fonológico. Estos resultados mostraron que, en un primer momento, el reconocimiento léxico de una palabra conlleva la activación de aquellas que presenten una forma fonológica similar; pero, una vez alcanzado el fonema desambiguador, esos competidores fonológicos pierden activación y son ahora los rasgos semánticos y visuales los que median en el reconocimiento léxico de la palabra *target* (Huettig y Altmann 2004; Dahan y Tanenhaus 2005).

Un último aspecto interesante acerca de los trabajos sobre activación léxica a partir del paradigma del mundo visual es el hecho de que una variable como la frecuencia léxica influya en el reconocimiento de la forma oral de las palabras, al igual que ocurría con su forma escrita. Dahan et al. (2001) contrastaron un elemento *target* (*bench* ‘banco’) con dos competidores fonológicos: uno de ellos con una mayor frecuencia léxica que el otro (*bed* ‘cama’ y *bell* ‘campana’, respectivamente). Cuando los participantes escuchaban los primeros fonemas (/be/), dirigían su mirada tanto al *target* como a sus competidores fonológicos; sin embargo, entre estos se registraba una diferencia significativa: el competidor fonológico con una mayor frecuencia léxica (*bed*) recibía una mayor proporción de fijaciones que aquel de menor frecuencia (*bell*).

En cuanto al procesamiento sintáctico, los primeros trabajos que emplearon el paradigma del mundo visual para examinar esta cuestión partieron, al igual que en los experimentos de lectura, de oraciones sintácticamente ambiguas. Tanenhaus et al. (1995) utilizaron como estímulo acústico estructuras en las que uno de sus constituyentes podía ser analizado como un complemento circunstancial o como un complemento del nombre, con el objetivo de examinar si, como en la lengua escrita, los

participantes caían en una vía muerta o si, por el contrario, se apoyaban en el contexto visual para solventar la ambigüedad. Simultáneamente, los participantes observaban una escena en la que aparecían cuatro objetos diferentes y, siguiendo las instrucciones del estímulo acústico, debían mover uno de ellos: *Put the apple on the towel in the box – pon la manzana encima de la toalla en la caja*–. Concretamente, la ambigüedad se situaba en el sintagma *on the towel*, pues se podía analizar como un complemento circunstancial –y, por tanto, el lugar de destino de la manzana– o como un complemento del nombre –en este caso, la manzana que estaba encima de la toalla (y no otra) era la que debía moverse a la caja–. Los resultados mostraron que en aquellas condiciones en las que solo se mostraba un referente del objeto que había que mover –en el ejemplo anterior, un solo referente de manzana–, había un incremento de las fijaciones hacia la imagen de la toalla cuando se escuchaba esta palabra (*towel*), indicativo ello de que los participantes entendían este constituyente como el complemento circunstancial – cayendo, por tanto, en una vía muerta–. Sin embargo, en el caso de mostrarse dos referentes del objeto que había que mover –una manzana encima de una toalla y otra encima de una servilleta–, los participantes dirigían una mayor proporción de sus fijaciones hacia la manzana que estaba encima de la toalla cuando escuchaban esta palabra (*towel*), por lo que ahora este constituyente se interpretaba como un complemento del nombre. A partir de estos resultados, Tanenhaus et al. señalaron que, en caso de disponer de información contextual adecuada, los oyentes se valen de ella para resolver una ambigüedad sintáctica (Sedivy et al. 1999).

Más allá de los procesos de resolución de ambigüedades, el paradigma del mundo visual también se ha empleado para analizar si el contexto lingüístico puede ayudar a predecir la estructura de una oración no ambigua y, por tanto, facilitar el procesamiento sintáctico. Por ejemplo, Altmann y Kamide (1999) examinaron los patrones de movimientos oculares mientras los participantes escuchaban oraciones como (12):

- (12) The boy will eat/move the cake
'El niño comerá/moverá la tarta'

En cuanto percibían el verbo *eat* ('comer'), se registraba un aumento de las fijaciones hacia el único objeto comestible de la escena (la tarta) en comparación con el verbo *move* ('mover'). Estos resultados indican que la información lingüística – concretamente, semántica– impone restricciones sobre la estructura de una oración, de tal forma que, antes de escuchar la palabra *target* –en el ejemplo anterior, *cake*–, los participantes son capaces de predecir este constituyente y dirigir su mirada hacia él. Conclusiones similares se han establecido en aquellos trabajos en los que se ha manipulado el orden canónico de los constituyentes y sus roles temáticos (Kaiser y Trueswell 2004; Knoeferle et al. 2005), así como el género gramatical (Lew-Williams y Fernald 2007).

4.2.3. Producción lingüística

Como señalamos antes, el paradigma del mundo visual se puede emplear igualmente para estudiar la producción de mensajes lingüísticos de forma oral: en este caso, los participantes observan una escena con el objetivo de nombrar las imágenes que en ella aparecen o de describir el evento que representan. Gracias a esta aplicación, se han

podido analizar, desde la órbita de la producción lingüística, los dos procesos cognitivos comentados previamente en relación a la comprensión.

Por un lado, diversos estudios basados en el paradigma del mundo visual se han interesado en analizar si el proceso de activación léxica durante la producción oral se puede reflejar en los patrones de movimientos oculares y cómo. Por ejemplo, Meyer et al. (1998) presentaron pares de objetos para que los participantes los nombraran mientras se registraban sus movimientos oculares. Los resultados mostraron que los hablantes dirigían su mirada hacia el objeto que iban a nombrar entre 900 y 1000 ms antes de comenzar a articular su nombre; en cambio, durante la propia articulación de la palabra desviaban su mirada hacia el siguiente objeto de la escena. Así pues, los hablantes parecen dirigir su mirada hacia el elemento *target* para seleccionar el lema de la palabra y activar su código fonológico y, una vez logrado ese objetivo, desvían su atención hacia el siguiente elemento, de tal forma que no es necesario que observen el *target* mientras articulan su nombre (Meyer y van der Meulen 2000).

No obstante, la duración de las fijaciones sobre el objeto *target* antes de ser nombrado puede variar en función de diversos rasgos. Por ejemplo, la tipicidad de un objeto determina durante cuánto tiempo se fija una imagen antes de ser nombrada: cuanto más común o prototípico sea un objeto, más cortas serán las fijaciones (Meyer et al. 1998). De manera similar, el número de palabras que existen para referirse a un mismo objeto condiciona la duración de sus fijaciones, ya que, en caso de poder emplear dos o más palabras, la selección léxica se dificulta y, en consecuencia, se incrementa la duración de las fijaciones (Griffin 2001). Por último, la presencia previa de una palabra con una forma fonológica similar a la del objeto nombrado facilita su activación léxica, por lo que la duración de las fijaciones sobre el objeto *target* disminuye (Meyer y van der Meulen 2000). Una última variable que es interesante mencionar por presentar un comportamiento diferente en la producción oral respecto de la lectura es la longitud: mientras que esta variable condiciona de manera significativa la duración de las fijaciones sobre una palabra escrita, no parece ocurrir lo mismo durante el nombrado, pues no se observan diferencias en la duración de las fijaciones sobre un objeto, independientemente de que la palabra para denominarlo sea corta o larga (Meyer et al. 2003).

Más allá de la activación léxica, el paradigma del mundo visual también se ha empleado para analizar cómo un hablante planifica la estructura de un enunciado. En este caso, los participantes suelen describir la acción representada en una escena con el objetivo de determinar si existe alguna relación entre la estructura oracional producida y el patrón de movimientos oculares registrado. A este respecto, Griffin y Block (2000) constataron que el orden de los constituyentes de una oración refleja fielmente la distribución de las fijaciones: aquel elemento de la imagen que se fija primero es el que ocupará la primera posición en la oración y así consecutivamente (Bock et al. 2003).

Más interesante, sin embargo, es el hecho de que la distribución de las fijaciones en una escena no es aleatoria, sino que está sujeta a la prominencia de los objetos que en ella aparezcan: cuanto más prominente sea un objeto, mayor es la probabilidad de que se fije en primer lugar y, en consecuencia, de que ocupe la primera posición en una oración, determinando así el tipo de estructura sintáctica que se construirá. Varios estudios han tratado de señalar qué factores definen la prominencia de un objeto y, entre ellos, el rasgo que probablemente haya recibido una mayor atención es la animacidad. Por

ejemplo, Esaulova et al. (2019) analizaron los movimientos oculares de los participantes mientras examinaban escenas en las que aparecía un agente, siempre animado, y un paciente, que podía ser animado o inanimado; al mismo tiempo, debían producir oraciones que describiesen la acción representada en la escena. Los resultados mostraron que la probabilidad de recibir una primera fijación, así como su duración, era mayor en los pacientes animados versus inanimados. Es decir, un ente animado y, sobre todo humano, tiende a ser más prominente y, por ende, a recibir las primeras fijaciones dentro de una escena (Humphreys et al. 2016). Este patrón de movimientos oculares tuvo, a su vez, consecuencias en el tipo de estructura sintáctica empleada para describir la escena: los pacientes animados recibían una mayor proporción de primeras fijaciones que los pacientes inanimados y, en consecuencia, tendían a ocupar la posición de sujeto dentro de la oración, lo que conducía a la producción de un mayor porcentaje de oraciones pasivas que de activas (Humphreys et al. 2016).

Pese a los resultados anteriores, algunos estudios cuestionan que la prominencia de los objetos y su consecuente activación léxica sean los únicos factores que determinen el tipo de estructura oracional que los hablantes construyen. Ciertos trabajos basados también en el paradigma del mundo visual parecen mostrar que la planificación de la estructura sintáctica puede ser anterior a la selección léxica. Por ejemplo, van der Meulen y Meyer (2000) presentaron cuatro objetos dentro de una escena –dos en la parte superior y dos en la parte inferior– y los hablantes debían indicar qué objetos estaban encima de cuáles. Los objetos de la parte inferior podían ser iguales, lo que conduciría a la producción de una oración simple –*la cuchara y la servilleta están encima de las tazas*–, o podían ser diferentes, dando lugar a dos oraciones coordinadas –*la cuchara está encima de la taza y la servilleta está encima de la jarra*–. En este estudio, antes de mirar el objeto que era articulado en primer lugar, los participantes dirigían la mirada hacia la zona inferior de la pantalla con el objetivo de determinar si debían articular una única oración o dos oraciones coordinadas. Este resultado sugiere, por tanto, que existe cierta planificación sintáctica previa a la activación léxica.

Dado el límite de espacio, no analizaremos la utilidad del registro de movimientos oculares para el estudio de otros procesos cognitivos involucrados, igualmente, en la comprensión y producción de mensajes orales. Sin embargo, sí es interesante señalar que, en los últimos años, los trabajos basados en el paradigma del mundo visual han intentado combinar estas dos órbitas con el objetivo de aplicar esta metodología al análisis del diálogo y de la interacción lingüística (Kreysa y Pickering 2011).

5. Conclusiones y perspectivas futuras

En el presente trabajo hemos intentado explicar, por un lado, en qué consiste la metodología *eye-tracking* y cómo se pueden registrar los movimientos de los ojos y, por otro, cuáles son las principales aportaciones de esta técnica al estudio del procesamiento del lenguaje. A este respecto, el registro de movimientos oculares se presenta como un método apropiado tanto para el estudio de la comprensión como de la producción lingüística, aunque hasta la fecha su uso en el campo de la comprensión, especialmente la escrita, ha generado un mayor número de investigaciones. No obstante, desde hace varios años, y gracias al surgimiento del paradigma del mundo visual, es posible

emplear esta metodología para el estudio de la comprensión y la producción de mensajes orales.

Independientemente de la órbita en la que nos movamos –comprensión o producción, lenguaje oral o escrito–, la metodología *eye-tracking* ha permitido analizar diversos procesos cognitivos involucrados en el correcto manejo del lenguaje, aunque en este trabajo nos hemos centrado, fundamentalmente, en dos: la activación léxica y la integración o la planificación de la estructura oracional. En particular, a través de los patrones de movimientos oculares se ha podido examinar la influencia que diversas variables lingüísticas como la frecuencia, la longitud o la ambigüedad ejercen sobre los procesos cognitivos de reconocimiento y selección del léxico o de integración en una estructura oracional.

Pese a las indudables contribuciones que esta metodología ha aportado al campo de la psicolingüística, todavía son varias las cuestiones que quedan abiertas y que, con suerte, podrán hallar respuesta en investigaciones futuras. Entre ellas, podemos destacar la necesidad de contemplar las diferencias individuales: no todos los hablantes presentan una misma idiosincrasia y algunos rasgos, como la capacidad de memoria operativa o la competencia lectora, pueden ejercer una influencia determinante sobre los patrones de movimientos oculares y, en consecuencia, sobre los procesos cognitivos involucrados en una tarea (Traxler et al. 2012; Bargary et al. 2017).

Un segundo reto de esta metodología –y, en general, de la disciplina psicolingüística– es poner en relación los resultados obtenidos hasta la fecha sobre movimientos oculares con los cosechados a partir de otros métodos. A este respecto, son varios los autores que se han interesado por combinar el registro de movimientos oculares con otras técnicas, como los potenciales evocados relacionados con eventos (Baccino 201) o la resonancia magnética funcional (Hanke et al. 2020); sin embargo, estos intentos son todavía escasos (para una introducción a ambas técnicas, véase Haro 2022 y Marrero-Aguilar 2022, respectivamente).

En último lugar, debemos destacar la necesidad de aplicar la metodología *eye-tracking* a tareas que permitan estudiar el procesamiento del lenguaje de una forma más natural. Por ejemplo, los estudios de lectura parten, generalmente, de oraciones aisladas y, aunque es cierto que esta forma de trabajar ha permitido conocer diversos aspectos sobre los procesos cognitivos involucrados en su comprensión, se vuelve necesario el empleo de materiales con una mayor validez ecológica, es decir, de materiales que simulen con mayor fidelidad las condiciones naturales de lectura (Spinner et al. 2013; Orquin y Holmqvist 2018).

Así pues, la metodología *eye-tracking* presenta todavía diversos retos a los que tendrá que enfrentarse en los próximos años, pero no por ello debemos dejar de destacar la gran aportación que este método ha hecho al campo de la psicolingüística para conocer algunos de los procesos cognitivos involucrados en una capacidad que nos hace tan humanos como es el lenguaje.

6. Referencias

- Allopenna, Paul D.; Magnuson, James S.; Tanenhaus, Michael K. 1998. Tracking the time course of spoken word recognition: Evidence for continuous mapping models. *Journal of Memory and Language* 38: 419-439.
- Altarriba, Jeanette; Kambe, Gretchen; Pollatsek, Alexander; Rayner, Keith. 2001. Semantic codes are not used in integrating information across eye fixations in reading: Evidence from fluent Spanish-English bilinguals. *Perception & Psychophysics* 63: 875-890.
- Altmann, Gerry T. M.; Kamide, Yuki. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73: 247-264.
- Altmann, Gerry T. M.; Garnham, Alan; Dennis, Yvette. 1992. Avoiding the garden path: Eye movements in context. *Journal of Memory and Language* 31.5: 685-712.
- Altmann, Gerry T. M.; Garnham, Alan; Henstra, Judith-Ann. 1994. Effects of syntax in human sentence parsing: Evidence against a structure-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20.1: 209-216.
- Baccino, Thierry 2011. Eye movements and concurrent event-related potentials: Eye fixation-related potential investigations in reading. En S. P. Liversedge, I. D. Gilchrist, y S. Everling, eds. *The Oxford handbook of eye movements*. Oxford: Oxford University Press, pp. 857-870.
- Balota, David A.; Pollatsek, Alexander; Rayner, Keith. 1985. The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology* 17.3: 364-390.
- Bargary, Gary; Bosten, Jenny M.; Goodbourn, Patrick T.; Lawrance-Owen, Adam J.; Hogg, Ruth E.; Mollon, John D. 2017. Individual differences in human eye movements: An oculomotor signature? *Vision Research* 141: 157-169
- Betancort, Moises; Carreriras, Manuel; Sturt, Patrick. 2009. The processing of subject and object relative clauses in Spanish: An eye-tracking study. *The Quarterly Journal of Experimental Psychology* 62.10: 1915-1929.
- Binder, Katherine S.; Duffy, Susan A.; Rayner, Keith. 2001. The effects of thematic fit and discourse context on syntactic ambiguity resolution. *Journal of Memory and Language* 44.2: 297-324.
- Bock, Kathryn; Irwin, David E.; Davidson, Douglas J.; Levelt, Willem J. M. 2003. Minding the clock. *Journal of Memory and Language* 48: 653-685.
- Boland, Julie E.; Blodgett, Allison. 2001. Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language* 45.3: 391-411.
- Calvo, Manuel G.; Meseguer, Enrique. 2002. Eye movements and processing stages in reading: Relative contribution of visual, lexical and contextual factors. *The Spanish Journal of Psychology* 5.1: 66-77.
- Carpenter, Patricia A.; Just, Marcel A. 1983. What your eyes do while your mind is reading. En K. Rayner, ed. *Eye movements in reading: Perceptual and language processes*. San Diego: Academic Press, pp. 275-307.
- Carpenter, Patricia A.; Daneman, Meredyth. 1981. Lexical retrieval and error recovery in reading: A model based on eye fixations. *Journal of Verbal Learning and Verbal Behavior* 28: 138-160.

- Carter, Benjamin T.; Luke, Steven G. 2020. Best practices in eye tracking research. *International Journal of Psychology* 155: 49-62.
- Clifton, Charles; Staub, Adrian; Rayner, Keith. 2007. Eye movements in reading words and sentences. En R. van Gompel, M. Fisher, W. Murray y R. L. Hill, eds. *Eye Movements: A Window of Mind and Brain*. Amsterdam: Elsevier, pp. 341-371.
- Clifton, Charles; Traxler, Matthew J.; Mohamed, Taha; Williams, Rihana S.; Morris, Robin K.; Rayner, Keith. 2003. The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language* 49: 317-334.
- Cooper, Roger M. 1974. The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology* 6: 84-107.
- Dahan, Delphine; Magnuson, James S.; Tanenhaus, Michael K. 2001. Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology* 42: 317-367.
- Dahan, Delphine; Tanenhaus, Michael K. 2005. Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychological Bulletin & Review* 12: 455-459.
- Dahan, Delphine; Tanenhaus, Michael, K.; Salverda, Anne P. 2007. The influence of visual processing on phonetically driven saccades in the “visual world” paradigm. En R. van Gompel, M. Fisher, W. Murray y R. L. Hill, eds. *Eye Movements: A Window of Mind and Brain*. Amsterdam: Elsevier, pp. 471-486.
- Drieghe, Denis. 2008. Foveal processing and word skipping during reading. *Psychonomic Bulletin & Review* 15: 856-860.
- Drieghe, Denis; Brysbaert, Marc; Desmet, Timothy; De Baecke, Constantijn. 2004. Word skipping in reading: On the interplay of linguistic and visual factors. *European Journal of Cognitive Psychology* 16.1-2: 79-103.
- Duffy, Susan A.; Keir, Jessica A. 2004. Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition* 32.4: 551-559.
- Duffy, Susan A.; Morris, Robin K.; Rayner, Keith. 1988. Lexical ambiguity and fixation times in reading. *Journal of Memory and Language* 27.4: 429-446.
- Duffy, Susan A.; Rayner, Keith. 1990. Eye Movements and Anaphor Resolution: Effects of Antecedent Typicality and Distance. *Language and Speech* 33.2: 103-119.
- Ehrlich, Susan F.; Rayner, Keith. 1981. Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning & Verbal Behavior* 20.6: 641-655.
- Ehrlich, Susan F.; Rayner, Keith. 1983. Pronoun assignment and semantic integration during reading: eye movements and immediacy of processing. *Journal of Verbal Learning & Verbal Behavior* 22.1: 75-87.
- Esaulova, Yulia; Penke, Martina; Dolscheid, Sarah. 2019. Describing events: Changes in eye movements and language production due to visual and conceptual properties of scenes. *Frontiers of Psychology*, 10. DOI: 10.3389/fpsyg.2019.00835
- Ferreira, Fernanda; Clifton, Charles. 1986. The independence of syntactic processing. *Journal of Memory and Language* 25: 348-368.

- Ferreira, Fernanda; Henderson, John M. 1990. Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16.4: 555-568.
- Fisher, Denis F.; Shebilske, Wayne L. 1985. There is more that meets the eye than the eyemind assumption. En R. Groner, G. W. McConkie, y C. Menz, eds. *Eye movements and human information processing*. Amsterdam: North Holland, pp. 149-158.
- Folk, Jocelyn R.; Morris, Robin K. 2003. Effects of syntactic category assignment on lexical ambiguity resolution in reading: An eye movement analysis. *Memory & Cognition* 31: 87-99.
- Frazier, Lyn. 1987. Sentence processing: A tutorial review. En M. Colheart, ed. *Attention and performance, XII: The psychology of reading*. Hillsdale: Lawrence Erlbaum Associates, pp. 559-586.
- Frazier, Lyn; Rayner, Keith. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14.2: 178-210.
- Frazier, Lyn; Rayner, Keith. 1987. Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language* 26: 505-526.
- Garnsey, Susan M.; Pearlmutter, Neal J.; Myers, Elizabeth; Lotocky, Melanie A. 1997. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language* 37.1: 58-93.
- Griffin, Zenzi M. 2001. Gaze durations during speech reflect word selection and phonological encoding. *Cognition* 82.1: B1-B14.
- Griffin, Zenzi M.; Bock, Kathryn. 2000. What the eyes say about speaking. *Psychological Science* 11: 274-279.
- Hanke, Michael; Mathôt, Sebastiaan; Ort, Eduard; Peitek, Norman; Stadler, Jörg; Wagner, Adina. 2020. A practical guide to functional magnetic resonance imaging with simultaneous eye tracking for cognitive neuroimaging research. En S. Pollmann, ed. *Spatial learning and attention guidance*. New York: Humana Press, pp. 291-305.
- Haro, Juan. 2022. Una introducción al uso de los potenciales evocados en el estudio del lenguaje. *Estudios de Lingüística del Español* 45: 185-204.
https://doi.org/10.1007/7657_2019_31
- Henderson, John M.; Dixon, Peter; Petersen, Alan; Twilley, Leslie C.; Ferreira, Fernanda. 1995. Evidence for the use of phonological representations during transsaccadic word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 21.1: 82-97.
- Henderson, John M.; Ferreira, Fernanda. 1993. Eye movement control during reading: Fixation measures reflect foveal but not parafoveal processing difficulty. *Canadian Journal of Experimental Psychology* 47.2: 201-221.
- Hendrickson, Anita E. 2009. Fovea: Primate. En L. R. Squire, ed. *Encyclopedia of Neuroscience*. Cambridge: Academic Press, pp. 327-334.
- Hoeks, John C.; Hendriks, Petra; Vonk, Wietske; Brown, Colin M.; Hagoort, Peter. 2006. Processing the noun phrase versus sentence coordination ambiguity: Thematic information does not completely eliminate the processing difficulty. *Quarterly Journal of Experimental Psychology* 59.9: 1581-1599.

- Hoeks, John C.; Vonk, Wietske; Schriefers, Herbert. 2002. Processing coordinated structures in context: The effect of topic-structure on ambiguity resolution. *Journal of Memory and Language* 46.1: 99-119.
- Holmes, Virginia M.; O'Regan, J. Kevin. 1981. Eye Fixations Patterns during the Reading of Relative-Clause Sentences. *Journal of Verbal Learning and Verbal Behavior* 20: 417-430.
- Hooge, Ignace T. C.; Holleman, Gijs A.; Haukes, Nina C.; Hessels, Roy S. 2019. Gaze tracking accuracy in humans: One eye is sometimes better than two. *Behavior Research Methods* 51.6: 2712-2721.
- Huettig, Falk; Altmann, Gerry T. M. 2004. The online processing of ambiguous and unambiguous words in context: Evidence from head mounted eye tracking. En M. Carreiras y C. Clifton, eds. *The on-line study of sentence comprehension: Eyetracking, ERP and beyond*. New York: Psychology Press, pp. 187-207.
- Huettig, Falk; McQueen, James M. 2007. The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language* 57.4: 460-482.
- Humphreys, Gina F.; Mirkovic, Jelena; Gennari, Silvia P. 2016. Similarity-based competition in relative clause production and comprehension. *Journal of Memory and Language* 89: 200-221.
- Hyönä, Jukka; Pollatsek, Alexander. 1998. Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance* 24.6: 1612-1627.
- Ikeda, Mitsuo; Saida, Shinya. 1978. Span of recognition in reading. *Vision Research* 18: 83-88.
- Inhoff, Albrecht W. 1984. Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning & Verbal Behavior* 23.5: 612-624.
- Inhoff, Albrecht W. 1989a. Parafoveal processing of words and saccade computation during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance* 15: 544-555.
- Inhoff, Albrecht W. 1989b. Lexical access during eye fixations in reading: Are word access codes used to integrate lexical information across interword fixations? *Journal of Memory and Language* 28: 444-461.
- Inhoff, Albrecht W.; Liu, Weimin. 1998. The perceptual span and oculomotor activity during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance* 24: 20-34.
- Inhoff, Albrecht W.; Pollatsek, Alexander; Posner, Michael I.; Rayner, Keith. 1989. Covert attention and eye movements during reading. *Quarterly Journal of Experimental Psychology* 41A: 63-89.
- Inhoff, Albrecht W.; Radach, Ralph; Eiter, Brianna M.; Juhasz, Barbara. 2003. Distinct subsystems for the parafoveal processing of spatial and linguistic information during eye fixations in reading. *Quarterly Journal of Experimental Psychology* 56A: 803-828.
- Inhoff, Albrecht W.; Rayner, Keith. 1986. Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics* 40: 431-439.
- Johnson, Marcus L.; Lowder, Matthew W.; Gordon, Peter C. 2011. The Sentence-Composition Effect: Processing of Complex Sentences Depends on the

- Configuration of Common Noun Phrase Versus Unusual Noun Phrases. *Journal of Experimental Psychology: General* 140.4: 707-724.
- Juhasz, Barbara J. 2005. Age-of-Acquisition Effects in Word and Picture Identification. *Psychological Bulletin* 131.5: 684-712.
- Juhasz, Barbara J. 2007. The influence of semantic transparency on eye movements during English compound word recognition. En M. J. Traxler y M. A. Gernsbacher, eds. *Handbook of Psycholinguistics*. Cambridge: Academic Press, pp. 373-389.
- Juhasz, Barbara J.; Rayner, Keith. 2003. Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29.6: 1312-1318.
- Juhasz, Barbara J.; Rayner, Keith. 2006. The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition* 13.7-8: 846-863.
- Juhasz, Barbara J.; Starr, Matthew; Inhoff, Albrecht W.; Placke, Lars. 2003. The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions, and eye fixations. *British Journal of Psychology* 94: 223-244.
- Just, Marcel A.; Carpenter, Patricia A. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review* 87: 329-354.
- Kaiser, Elsi; Trueswell, John C. 2004. The role of discourse context in the processing of a flexible word-order language. *Cognition* 94: 113-147.
- Knoeferle, Pia; Crocker, Matthew W.; Scheepers, Christoph; Pickering, Martin J. 2005. The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition* 95: 95-127.
- Kolb, Helga; Fernández, Eduardo; Nelson, Ralph. 1995. *Webvision: The organization of the retina and visual system*. Salt Lake City: University of Utah Health Sciences Center.
- Kreysa, Helene; Pickering, Martin J. 2011. Eye movements in dialogue. En S. P. Liversedge, I. D. Gilchrist y S. Everling, eds. *The Oxford handbook of eye movements*. Oxford: Oxford University Press, pp. 943-959.
- Lew-Williams, Casey; Fernald, Anne. 2007. Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological science* 18.3: 193-198.
- Lima, Susan D. 1987. Morphological analysis in sentence reading. *Journal of Memory and Language* 26: 84-99.
- Liversedge, Simon P.; Rayner, Keith; White, Sarah J.; Findlay, John M.; McSorley, Eugene. 2006. Binocular coordination of the eyes during reading. *Current Biology* 16: 1726-1729.
- Liversedge, Simon P.; White, Sarah J.; Findlay, John M.; Rayner, Keith. 2006. The binocular coordination of eye movements during reading in children and adults. *Vision Research* 46: 3898-3908.
- MacDonald, Maryellen C. 1997. Lexical Representations and Sentence Processing: An Introduction. *Language and Cognitive Processes* 12.2/3: 121-136.
- Mak, Willem M.; Vonk, Wietske; Schriefers, Herbert. 2002. The Influence of Animacy on Relative Clause Processing. *Journal of Memory and Language* 47: 50-68.

- Mak, Willem M.; Vonk, Wietske; Schriefers, Herbert. 2006. Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language* 54: 466-490.
- Marrero-Aguiar, Victoria. 2022. Aportación de las investigaciones con neuroimagen funcional a los estudios lingüísticos: algunas preguntas y respuestas. *Estudios de Lingüística del Español* 45: 159-183.
- Matin, Ethel. 1974. Saccadic suppression: A review. *Psychological Bulletin* 81: 899-917.
- McConkie, George W.; Kerr, Paul W.; Reddix, Michael D.; Zola, David; Jacobs, Arthur M. 1989. Eye movement control in reading: II. Frequency of refixating a word. *Perception & Psychophysics* 46: 245-253.
- McConkie, George W.; Rayner, Keith. 1975. The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics* 17: 578-586.
- McConkie, George W.; Rayner, Keith. 1976. Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society* 8: 365-368.
- Meyer, Antjia S.; Roelofs, Ardi; Levelt, Willem J. M. 2003. Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language* 48.1: 131-147.
- Meyer, Antjia S., Sleiderink, Astrid M.; Levelt, Willem J. M. 1998. Viewing and naming objects. *Cognition* 66: B25-B33.
- Meyer, Antjia S.; van der Meulen, Femke F. 2000. Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review* 7: 314-319.
- O'Regan, J. Kevin; Lévy-Schoen, Ariane. 1987. Eye movement strategy and tactics in word recognition and reading. En M. Coltheart, ed. *Attention and performance XII. The psychology of reading*. Hillsdale: Erlbaum pp. 363-383.
- Orquin, Jacob, L.; Holmqvist, Kenneth. 2018. Threats to the validity of eye-movement research in psychology. *Behavior Research Methods* 50: 1645-1656.
- Osaka, Naoyuki. 1987. Effect of peripheral visual field size upon eye movements during Japanese text processing. En J. K. O'Regan y A. Levy-Schoen, eds. *Eye movements: From physiology to cognition*. Amsterdam: North Holland, pp. 421-429.
- Osaka, Naoyuki; Oda, Koichi. 1991. Effective visual field size necessary for vertical reading during Japanese text processing. *Bulletin of the Psychonomic Society* 29: 345-347.
- Pickering, Martin J.; Traxler, Matthew J.; Crocker, Matthew W. 2000. Ambiguity Resolution in Sentence Processing: Evidence against Frequency-Based Accounts. *Journal of Memory and Language* 43: 447-475.
- Pollatsek, Alexander; Bolozky, Shmuel; Well, Arnold D.; Rayner Keith. 1981. Asymmetries in the perceptual span for Israeli readers. *Brain and Language* 14: 174-180.
- Pollatsek, Alexander; Hyönä, Jukka; Bertram, Raymond. 2000. The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance* 26.2: 820-833.
- Pollatsek, Alexander; Lesch, Mary; Morris, Robin K.; Rayner, Keith. 1992. Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance* 18.1: 148-162.

- Pollatsek, Alexander; Rayner, Keith; Balota, David A. 1986. Inferences about eye movement control from the perceptual span in reading. *Perception & Psychophysics* 40: 123-130.
- Raney, Gary E.; Cambell, Spencer J.; Bovee, Joanna C. 2014. Using eye movements to evaluate the cognitive processes involved in text comprehension. *Journal of Visualized Experiments* 83: 1-7.
- Rayner, Keith. 1975. The perceptual span and peripheral cues in reading. *Cognitive Psychology* 7: 65-81.
- Rayner, Keith. 1979. Eye guidance in reading: Fixation locations within words. *Perception* 8: 21-30.
- Rayner, Keith. 1986. Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology* 41: 211-236.
- Rayner, Keith. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124: 372-422.
- Rayner, Keith. 2009. Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology* 62: 1457-1506.
- Rayner, Keith; Balota, David A.; Pollatsek, Alexander. 1986. Against parafoveal semantic preprocessing during eye fixations in reading. *Canadian Journal of Psychology* 40: 473-483.
- Rayner, Keith; Bertera, James H. 1979. Reading without a fovea. *Science* 206: 468-469.
- Rayner, Keith; Carlson, Marcia; Frazier, Lyn. 1983. The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning & Verbal Behavior* 22.3: 358-374.
- Rayner, Keith; Duffy, Susan A. 1986. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition* 14: 191-201.
- Rayner, Keith; Liversedge, Simon P. 2011. Linguistic and cognitive influences on eye movements during reading. En S. P. Liversedge, I. D. Gilchrist y S. Everling, eds. *The Oxford handbook of eye movements*. Oxford: Oxford University Press, pp. 751-766.
- Rayner, Keith; Pollatsek, Alexander. 1989. *The psychology of reading*. Englewood Cliffs: Prentice Hall.
- Rayner, Keith; Pollatsek, Alexander. 2007. Eye-movement control in reading. En M. J. Traxler y M. A. Gernsbacher, eds. *Handbook of Psycholinguistics*. Cambridge: Academic Press, pp. 613-657.
- Rayner, Keith; Raney, Gary E.; Pollatsek, Alexander. 1995. Eye movements and discourse processing. En R. F. Lorch y E. J. O'Brien, eds. *Sources of coherence in reading*. Hillsdale: Erlbaum, pp. 9-36.
- Rayner, Keith; Reichle, Erik D.; Stroud, Michael J.; Williams, Carrick C.; Pollatsek, Alexander. 2006. The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging* 21.3: 448-465.
- Rayner, Keith; Sereno, Sara C.; Morris, Robin K.; Schmauder, A. Réne; Clifton, Charles. 1989. Eye movements and on-line language comprehension processes. *Language and Cognitive Processes* 4.3-4: SI21-SI49.

- Rayner, Keith; Sereno, Sara C.; Raney, Gary E. 1996. Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance* 22.5: 1188-1200.
- Rayner, Keith; Warren, Tessa; Juhasz, Barbara J.; Liversedge, Simon P. 2004. The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition* 30: 1290-1301.
- Rayner, Keith; Well, Arnold D.; Pollatsek, Alexander. 1980. Asymmetry of the effective visual field in reading. *Perception & Psychophysics* 27: 537-544.
- Rayner, Keith; Well, Arnold D.; Pollatsek, Alexander; Bertera, James H. 1982. The availability of useful information to the right of fixation in reading. *Perception & Psychophysics* 31: 537-550.
- Rayner, Keith; White, Sarah J.; Kambe, Gretchen; Miller, Brett; Liversedge, Simon P. 2003. On the processing of meaning from parafoveal vision during eye fixations in reading. En J. Hyönä, R. Radach y H. Deubel, eds. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*. Amsterdam: North Holland, pp. 213-234.
- Salverda, Anne P.; Tanenhaus, Michael, K. 2017. The visual world paradigm. En A. M. B. de Groot y P. Hagoort, eds. *Research methods in psycholinguistics and the neurobiology of language: A practical guide*. New York: Wiley Blackwell, pp. 89-110.
- Sedivy, Julie C.; Tanenhaus, Michael K.; Chambers, Craig G.; Carlson, Gregory N. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition* 71.2: 109-147.
- Spinner, Patti; Gass, Susan M.; Behney, Jennifer. 2013. Ecological validity in eye-tracking: An empirical study. *Studies in Second Language Acquisition* 35.2: 389-415.
- Spivey, Michael J.; Tanenhaus, Michael K. 1998. Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24.6: 1521-1543.
- Staub, Adrian. 2010. Eye movements and processing difficulty in object relative clauses. *Cognition* 116: 71-86.
- Sturt, Patrick. 2003. The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language* 48.3: 542-562.
- Swingley Daniel; Pinto John P.; Fernald Anne. 1999. Continuous processing in word recognition at 24 months. *Cognition* 71.2: 73-108.
- Tanenhaus, Michael, K. 2007. Eye movements and spoken language processing. En R. van Gompel, M. Fisher, W. Murray y R. L. Hill, eds. *Eye Movements: A Window of Mind and Brain*. Amsterdam: Elsevier, pp. 443-469.
- Tanenhaus Michael K.; Spivey-Knowlton Michael J.; Eberhard Kathleen M.; Sedivy Julie C. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268: 1632-1634.
- Tanenhaus, Michael K.; Trueswell, John C. 2006. Eye movements and spoken language comprehension. En M. J. Traxler y M. A. Gernsbacher, eds. *Handbook of Psycholinguistics*. Cambridge: Academic Press, pp. 863-900.
- Traxler, Matthew J.; Long, Debra L.; Tooley, Kristen M.; Johns, Clinton L.; Zirnstein, Megan; Jonathan, Eunike. 2012. Individual differences in eye-movements during

- reading: Working memory and speed-of-processing effects. *Journal of eye movement research* 5.1: 5.
- Traxler, Matthew J.; Morris, Robin K.; Seely, Rachel E. 2002. Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language* 47: 69-90.
- Traxler, Matthew J.; Williams, Rihana S., Blozis, Shelley A.; Morris, Robin K. 2005. Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language* 53: 204-224.
- Trueswell, John C.; Tanenhaus, Michael K.; Garnsey, Susan M. 1994. Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33.3: 285-318.
- Trueswell, John C.; Tanenhaus, Michael K.; Kello, Christopher. 1993. Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19.3: 528-553.
- Underwood, N. Roderick; McConkie, George W. 1985. Perceptual span for letter distinctions during reading. *Reading Research Quarterly* 20: 153-162.
- van der Meulen, Femke F.; Meyer, AS. 2000. Coordination of eye gaze and speech in sentence production. Póster presentado en *41st Annual Meeting of the Psychonomic Society, noviembre 16-19, 2000*. Nueva Orleans, LA.
- Vasishth, Shravan; von der Malsburg, Titus; Engelmann, Felix. 2013. What eye movements can tell us about sentence comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science* 4.2: 125-134.
- Vitu, Françoise; McConkie, George W.; Zola, David. 1998. About regressive saccades in reading and their relation to word identification. En G. Underwood, ed. *Eye guidance in reading and scene perception*. Oxford: Elsevier Science, pp. 101-124.
- Wade, Nicholas J.; Tatler, Benjamin W. 2005. *The moving tablet of the eye. The origins of modern eye movement research*. Oxford: Oxford University Press.
- Warren, Tessa; McConnell, Kerry; Rayner, Keith. 2008. Effects of context on eye movements when reading about possible and impossible events. *Journal of experimental psychology. Learning, memory, and cognition* 34.4: 1001-1010.
- White, Sarah J.; Drieghe, Denis; Liversedge, Simon P.; Staub, Adrian. 2018. The word frequency effect during sentence reading: A linear or nonlinear effect of log frequency? *The Quarterly Journal of Experimental Psychology* 71.1: 46-55.
- White, Sarah J.; Rayner, Keith; Liversedge, Simon P. 2005a. The influence of parafoveal word length and contextual constraint on fixation durations and word skipping in reading. *Psychonomic Bulletin & Review* 12: 466-471.
- White Sarah J; Rayner Keith; Liversedge Simon P. 2005b. Eye movements and the modulation of parafoveal processing by foveal processing difficulty: A re-examination. *Psychonomic Bulletin & Review* 12.5: 891-896.
- Wilson, Michael P.; Garnsey, Susan M. 2009. Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language* 60: 368-392.
- Wolverton, Gary S.; Zola, David. 1983. The temporal characteristics of visual information extraction during reading. En K. Rayner, ed. *Eye movements in reading: Perceptual and language processes*. New York: Academic Press, pp. 41-52.

¹ Aunque estas son las principales medidas de movimientos oculares (y, por tanto, en las que nos centraremos en el presente trabajo), se pueden analizar otros patrones como el diámetro de la pupila (pupilometría) o movimientos diferentes a las sacadas (vergencia, movimiento vestibular, nistagmo...). Para un repaso de los diversos patrones de movimientos oculares, véase Rayner (1998) o Carter y Luke (2020).

² Esta práctica, sin embargo, genera cierto debate, pues se ha comprobado que, durante la lectura, el ojo izquierdo y el derecho están fijados sobre caracteres diferentes en el 40-50 % de las ocasiones (Liversedge, Rayner et al. 2006; Liversedge, White et al. 2006).

³ Hay que señalar que las medidas tempranas y tardías de movimientos oculares no se corresponden fielmente con las fases inicial y tardía de procesamiento que reconocen algunos modelos seriales de procesamiento lingüístico, como el modelo de vía muerta (Frazier y Rayner 1982; Frazier 1987). De hecho, suele existir una alta correlación entre las medidas tempranas y tardías de movimientos oculares, de tal forma que, si un efecto es significativo en una medida temprana, probablemente también lo será en una tardía (MacDonald 1997; Vasishth et al. 2013).

⁴ Si la región está compuesta por una sola palabra, se emplea el término *gaze duration* (Rayner y Duffy 1986); si consta de más de una palabra, se prefiere el nombre *first-pass reading time* (Pickering et al. 2000).

⁵ Existe cierta controversia acerca de si la medida de cuasi primer barrido, así como de tiempo de lectura con regresiones y probabilidad de regresión, debe considerarse como temprana o tardía (Clifton et al. 2007; Rayner y Liversedge 2011). Por un lado, esta medida incluye la duración de las primeras fijaciones, por lo que registra ese primer contacto con el texto, pudiendo ser considerada como temprana. No obstante, esta medida (así como las dos siguientes) también incluye las refijaciones en una palabra o región, pudiendo ser considerada por ello tardía. Dada esta controversia, algunos autores prefieren referirse a las medidas anteriores como “intermedias” y, de hecho, es el término que emplearemos en el presente trabajo.

⁶ A este respecto, Ferreira y Clifton (1986) introdujeron una práctica en aquellos casos en los que la longitud de las palabras que se contrastaban entre condiciones no pudiese ser equivalente. Concretamente, estos autores dividieron los tiempos brutos de lectura de cada palabra o región por su número de caracteres, incluyendo los espacios; el resultado de esta división daría lugar al tiempo de lectura por carácter. Pese al relativo éxito que esta práctica tuvo en un principio (Frazier y Rayner 1987; Garnsey et al. 1997; Boland y Blodgett 2001), hoy en día se recomienda evitarla, ya que asume una relación lineal que no siempre es real entre el tiempo de lectura y el número de caracteres (Trueswell et al. 1994; Rayner 1998).

⁷ Una oración de relativo reducida es aquella en la que el pronombre relativo y el verbo auxiliar se eliden. Compárese las oraciones relativas reducidas de (6) con su forma no reducida (y, en consecuencia, no ambigua): *The defendant that was examined by the lawyer turned out to be unreliable* o *The evidence that was examined by the lawyer turned out to be unreliable*.

⁸ Estos estudios suelen emplear, generalmente, dos paradigmas: el paradigma de la frontera (o *boundary paradigm*), que consiste en presentar una palabra *target* en la región parafoveal y, cuando la mirada cruza una frontera invisible y se posa sobre ella, se sustituye (o no) por otra (Rayner 1975, 1998, 2009); o el paradigma de la ventana móvil (o *moving window paradigm*), en el que una palabra *target* aparece cubierta completa o parcialmente por otras letras cuando se muestra en la parafovea y, cuando se fija, se cambia por su forma correcta (McConkie y Rayner 1975; Rayner 1998, 2009).

El *big data* en los estudios del lenguaje

Javier Valenzuela
Universidad de Murcia
jvalen@um.es

Resumen

El presente trabajo examina las posibilidades que los acercamientos basados en los *big data* ofrecen a la investigación sobre el lenguaje. De manera resumida, los *big data* o “macrodatos” son los datos masivos que los usuarios generan en sus interacciones con el mundo digital y cuyo ingente volumen y naturaleza heterogénea precisa de un tratamiento especializado. El trabajo revisa de manera inicial las principales características de los *big data* para centrarse a continuación en los posibles problemas derivados del uso de *big data* en los análisis lingüísticos. La siguiente sección ofrece una revisión de estudios concretos que utilizan este acercamiento aplicándolo a la multimodalidad: un estudio del lenguaje que incluye no sólo el componente verbal sino aspectos multimodales como la gestualidad o la entonación. El trabajo concluye con una revisión de las ventajas y los problemas de la utilización de este tipo de datos.

Palabras clave: lingüística, análisis de corpus, *big data*, multimodalidad.

Abstract

This paper examines the possibilities that big data-based approaches offer to language research. In a nutshell, the term “big data” makes reference to the massive amount of data that users generate in their digital interactions and whose great volume and heterogeneous nature typically requires a specialized treatment. The chapter starts by reviewing the main characteristics of big data, and then focuses on the possible problems arising from the use of big data in linguistic analysis. The following section offers a review of specific studies that apply this big-data approach to the study of multimodality: an approach to language study that includes not only the verbal component but also multimodal aspects such as gestures or intonation. The paper concludes with a review of the advantages and problems of using this type of data.

Keywords: linguistics, corpus analysis, big data, multimodality.

1. Introducción

El término *big data*, traducido a veces como “macrodatos” o “datos masivos”, aparece mencionado con gran frecuencia en los medios de comunicación; su uso está extendido por multitud de ámbitos que afectan nuestro día a día, a pesar de lo cual existe un cierto nivel de desconocimiento sobre su naturaleza y su tratamiento. Básicamente, se denomina *big data* a toda la información que se genera al utilizar los distintos dispositivos digitales (internet, teléfonos, tarjetas de crédito, etc.). Esta información es recopilada y utilizada por empresas (o incluso por agencias gubernamentales), que basan sus decisiones en el conocimiento aportado por este tipo de datos personales. Dada la extensión de su uso y

su gran efectividad para determinadas predicciones, la pregunta obvia es hasta qué punto son útiles estos datos masivos para actividades de corte más académico, y más específicamente, cómo de valiosos son para los estudios del lenguaje. Este trabajo revisa el concepto y el tratamiento de los *big data*, discute su aplicación al ámbito de los estudios lingüísticos y finaliza con una descripción de los acercamientos que intentan aplicar de la manera más efectiva posible las principales características de los *big data* a las ciencias lingüísticas. Para ello nos centramos en los acercamientos que abordan el estudio del lenguaje desde el punto de vista de la multimodalidad, es decir, entendiendo la comunicación lingüística en toda su complejidad, prestando atención no solo a las palabras que decimos sino a su entonación, a los gestos que realizamos, a la información transmitida con los ojos y la mirada, las expresiones faciales o en general cualquier tipo de información física presente en el acto comunicativo. El trabajo discute algunas de las oportunidades así como de los desafíos inherentes a la aplicación del concepto de *big data* a los estudios del lenguaje.

2. *Big data*: algunos conceptos básicos

Sin lugar a dudas, el anglicismo *big data*, traducido aproximadamente como “grandes datos”, “macrodatos” o “datos masivos”, es uno de los términos más usados, buscados y estudiados del momento. Una búsqueda en Google devuelve al menos 120 millones de resultados. Al parecer, estamos en la era del *big data*, que cobra una importancia cada vez mayor en nuestras vidas. Pero ¿qué es exactamente el *big data*?

El término *big data* hace referencia al conjunto de datos que generamos en nuestra interacción con el mundo digital, que claramente abarca un porcentaje amplísimo de nuestra actividad cotidiana. Es nuestro uso de los distintos dispositivos electrónicos modernos (lo que podemos denominar nuestra “interacción digital”) lo que sirve como fuente de estos datos. Así, todos los aparatos electrónicos conectados a la red recopilan información puntual y precisa sobre el uso que hacemos de ellos. Por ejemplo, los servidores de correo electrónico recogen estadísticas de cuáles son los momentos en los que escribimos un correo (o un número mayor de ellos), a quién le escribimos, sobre qué temas escribimos (el asunto de la privacidad es un tema espinoso, que trataremos más abajo) y qué programa utilizamos para ello. Pero no solo nuestros correos electrónicos: nuestras búsquedas en la red (por ejemplo, en Google) quedan registradas y generan información sobre cuándo y qué buscamos (consultas de opiniones sobre un producto, una película, un restaurante, un hotel, un concierto, un libro, la búsqueda de un determinado tema, etc., información que luego sirve para devolvernos publicidad dirigida a nuestros intereses específicos), desde qué dispositivo –móvil, tableta, ordenador, portátil o incluso en qué lugar estamos al realizar la búsqueda–; el uso del GPS en nuestros móviles (el uso de Google Maps, por ejemplo) o en los navegadores de nuestros coches genera información detallada espacial y temporal sobre nuestros desplazamientos; el pago con tarjeta de crédito informa de cuánto hemos gastado, cuándo, dónde y en qué; el pago de facturas informa sobre nuestro consumo y nivel de gasto, así como las transacciones bancarias por internet; las llamadas de teléfono informan de a qué horas llamamos con mayor frecuencia, cuánto duran esas llamadas o con quién nos comunicamos; qué películas vemos en Netflix, HBO, AmazonPrime, Movistar+, Disney+, Apple TV... La lista es realmente casi interminable. Y, por supuesto, no podemos olvidar nuestro uso de las redes sociales. Las redes sociales son un fenómeno absolutamente desbordante, que

nos permite contactar con otros usuarios para comunicarnos con ellos, informarles de nuestros gustos, hacer amistades, compartir experiencias, ligar, jugar, encontrar trabajo, enterarnos de opiniones de un tema, escuchar música, buscar alojamiento, compartir aficiones... Las redes sociales están por todos sitios y su número sigue creciendo, tanto en diversidad como en número de usuarios. Algunas de ellas son archiconocidas y omnipresentes (Facebook tiene 2800 millones de usuarios; recordemos que la población mundial se estima en 7500 millones de habitantes), otras redes son más especializadas. Además de Facebook podemos mencionar WhatsApp, YouTube, Twitter, Instagram, TikTok, Telegram, Vimeo, LinkedIn, Pinterest, Tuenti, Snapchat, Tinder, Yelp, Google Maps, Waze, WeChat, Skype, Fortnite, World of Warcraft, Tumblr, Reddit, Viber, Line, Spotify, Google +, Flickr, Slideshare, Soundcloud, Badoo, TripAdvisor, AirBnB, Foursquare, Quora... además de las usadas en China y menos explotadas en Occidente. Nuestro uso de estas redes sociales también genera una gran cantidad de información, específica de cada caso. El volumen de datos que se genera es tan absolutamente ingente que resulta difícil de conceptualizar, de ahí la denominación de esta información como *big data*. Este volumen de información supera la capacidad del *software* convencional para poder capturarla, administrarla o procesarla en un tiempo razonable. De todas maneras, lo que hace especialmente complicado el tratamiento de estos datos no es solo su volumen, sino su velocidad de crecimiento (se generan datos de manera continua a cada momento, véase Figura 1) así como su heterogeneidad (hay datos de muy distinta clase). Es por ello por lo que los estudios sobre *big data* suelen caracterizar estos datos con una serie de “V”: Volumen, Velocidad, Variedad, Visualización, Valor, Viralidad y Veracidad.



Figura 1. Datos generados en un minuto en el universo digital (<https://es.statista.com/>)

Y todo esto ocurre cuando todavía no ha llegado “El Internet de las Cosas”, que conectará a la red de internet los objetos con los que interactuamos diariamente. Esta conexión convertirá nuestras casas en hogares inteligentes (la famosa nevera que te avisa sobre el estado de sus alimentos o la falta de algunos, pero también termostatos inteligentes que permitan optimizar el consumo de energía, o avisos sobre tus medicamentos caducados). Este nuevo estadio de información implica que ya no habrá nunca más objetos fuera de *stock* al ir a comprar; tampoco se perderán las cosas (imaginemos las llaves del coche o

de casa conectadas a internet y geolocalizadas en todo momento). Las aplicaciones de esta revolución en marcha se extienden a prácticamente todos los ámbitos de nuestra vida: la medicina y la salud, el transporte, los procesos industriales, el consumo y la economía de mercado y un larguísimo etcétera que probablemente no somos capaces de imaginar aún. Obviamente, esta revolución multiplicará las posibilidades (y los problemas) de los *big data* en un grado difícil de prever. Un cálculo conservador cifraba en 26000 millones los dispositivos conectados en 2020. A las enormes dificultades derivadas de la gestión de un tamaño tan ingente de datos se suma la variedad de estos. Aunque es complicado establecer categorías fijas, de manera aproximada los datos recogidos se pueden clasificar en tres tipos:

(i) Datos estructurados: son datos que tienen un formato y una estructura fija, como la que podríamos encontrar en una base de datos. Estos datos son fácilmente explotables, puesto que es sencillo encontrar la información buscada (véase a la Figura 2):

	nombre	color	edad	altura	peso	puntuacion
1:	Paco	Rojo	24	182	74.8	83
2:	Juan	Green	30	170	70.1	500
3:	Andres	Amarillo	41	169	60.0	20
4:	Natalia	Green	22	183	75.0	865
5:	Vanesa	Verde	31	178	83.9	221
6:	Miriam	Rojo	35	172	76.2	413
7:	Juan	Amarillo	22	164	68.0	902

Figura 2. Datos estructurados

(ii) Datos no estructurados: estos datos no responden a un patrón concreto; pueden ser los movimientos de ratón en una página web, el recorrido captado por el GPS en nuestros desplazamientos, datos de audio con grabaciones de voz, incluso los datos textuales (comentarios escritos por los usuarios en las redes sociales), etc. tal y como aparece en la Figura 3. Claramente, los métodos de recuperación de la información que pueden aplicarse a estos datos son mucho más complicados; cada categoría dentro de ellos requiere de un procedimiento distinto de tratamiento y de procesado para poder extraer la información requerida.



Figura 3. Ejemplos de datos no estructurados (ratón, audio, texto y GPS)

(iii) Datos semiestructurados: estos datos combinan una estructura más o menos fija con datos abiertos (véase Figura 4).

```
{
  "marcadores": [
    {
      "latitude": 40.416875,
      "longitude": -3.703308,
      "city": "Madrid",
      "description": "Puerta del Sol"
    },
    {
      "latitude": 40.417438,
      "longitude": -3.693363,
      "city": "Madrid",
      "description": "Paseo del Prado"
    },
    {
      "latitude": 40.407015,
      "longitude": -3.691163,
      "city": "Madrid",
      "description": "Estación de Atocha"
    }
  ]
}
```

Figura 4. Ejemplo de datos semiestructurados

De todas las características de los *big data*, dos de ellas, su gran volumen y su variedad, son las principales responsables de la dificultad tanto de su almacenamiento como de su tratamiento y gestión (son precisamente estas dos características las más interesantes para los estudios de la lengua, como veremos más adelante). De esta manera, han surgido compañías que ofrecen esos servicios especializados (por ejemplo, Spark o Hadoop) tanto para el almacenamiento de estos datos (que suele realizarse en “granjas de servidores”, lugares especializados que reúnen multitud de ordenadores que son capaces de almacenar enormes cantidades de información), como para su tratamiento.

El tratamiento de los datos merece también un comentario especial. Han surgido una gran cantidad de técnicas de explotación dedicadas a su procesamiento. Suelen ser acercamientos estadísticos, que detectan patrones generales subyacentes a esa gran cantidad de datos que son muy difíciles (por no decir imposibles) de observar a simple vista. Proliferan los cursos especializados en este tipo de procesamiento, que aumenta su paleta de herramientas de manera progresiva: sin ninguna intención de ser exhaustivo, se pueden mencionar los análisis de regresión, análisis de series temporales, A/B Testing (*split testing*), algoritmos genéticos o aprendizaje por máquina (*machine learning*). Más relacionado con el lenguaje está también el “Análisis de Sentimientos” (*Sentiment Analysis*), que permite extraer automáticamente la opinión de los consumidores sobre un producto (una película, un electrodoméstico, un restaurante...) de sus comentarios en las redes sociales. Ahora bien, en realidad, la gran mayoría de los datos masivos se procesan utilizando los modelos conocidos como “redes neuronales” artificiales (también conocidas como “redes conexionistas”). De hecho, para muchos autores, esta nueva ola de datos masivos es la que ha permitido el gran auge de las redes neuronales, cuyo uso se inició en los años 80 (culminando en la obra clásica de Rumelhart y McClelland 1986), y que han alcanzado su mayor grado de extensión y desarrollo en estos momentos. Las redes neuronales son programas de ordenador que utilizan una arquitectura distinta de los programas más clásicos o “algorítmicos”. En estos programas algorítmicos, la información se trata de manera secuencial, como se hace típicamente en lenguajes de programación como Python, C++, Java o Javascript. De esta manera, la información va siguiendo una serie de “pasos” secuenciales, con determinados saltos condicionales (“si se cumple esta condición, haz tal cosa; si no, sigue por la siguiente instrucción”). Sin embargo, las redes neuronales funcionan de manera muy distinta: estos programas están formados por una serie de “nodos”, cuyo funcionamiento está inspirado en el comportamiento de las neuronas del cerebro. De esta manera, cada nodo puede recibir distintos grados de “activación”; esta activación se pasa en paralelo a todos los otros nodos con los que esté conectado. Existen distintas “capas” de nodos; en las redes multicapa, representadas en la Figura 5, al menos existe una capa de entrada (*input*), una capa de salida o respuesta y una o más capas “intermedias”. Uno de los modos de funcionamiento más populares requiere dos estadios. En el primero, o fase de entrenamiento, las redes intentan conectar, por medio de un proceso de prueba y error, un estado de entrada (el *input*) a una respuesta deseada (el *output*) y lo hacen modificando poco a poco la fuerza con la que un nodo se conecta al siguiente (lo que se conoce como los “pesos” de la red), hasta que se alcanza el resultado deseado. Una vez la red ha sido “entrenada” y consigue el resultado deseado con el set de datos de entrenamiento, esa red ha “aprendido” y es capaz de encontrar la solución correcta a casos nuevos, que no formaban parte de su set de entrenamiento, es decir, es capaz de generalizar, que es lo que hace tan útiles y poderosas a estas redes. Por ejemplo, se alimenta una red con una serie de fotos de lunares (que sería el *input*), unos de ellos cancerígenos y otros benignos

(esta clasificación sería el *output* buscado). Una vez que la red ha sido entrenada de manera adecuada, y ha alcanzado la configuración de pesos o conexiones que permite distinguir entre ambos tipos de lunares, se le pueden dar una nueva serie de datos (lunares que no ha visto nunca) y será capaz de predecir con un altísimo nivel de confianza si los lunares son cancerígenos o no.

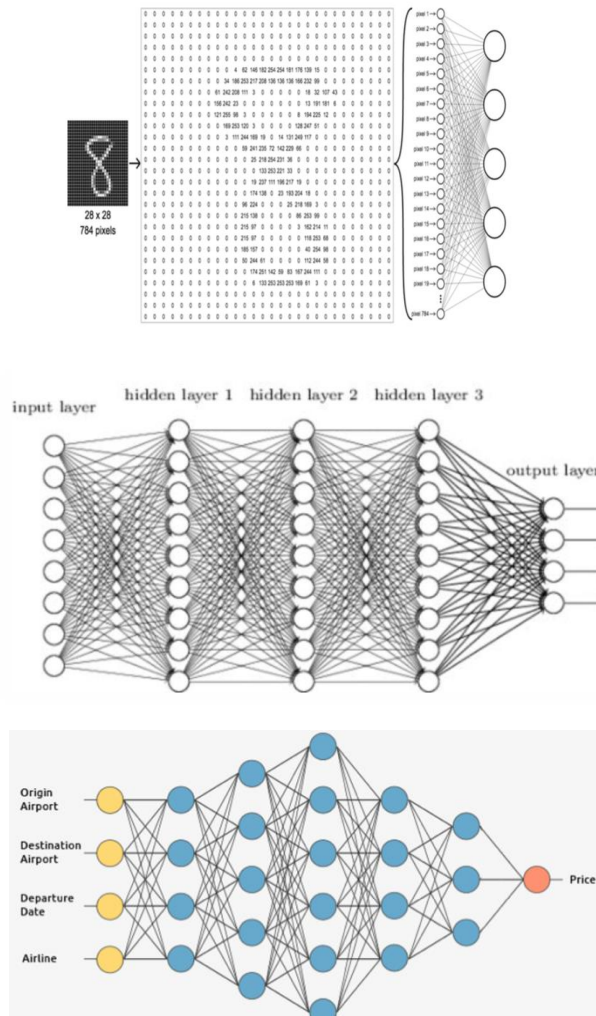


Figura 5. Ejemplos de redes neuronales multicapa

Las redes neuronales son herramientas muy efectivas; una vez entrenadas con una gran cantidad de datos, son capaces de generalizar con un alto grado de éxito. Sin embargo, tienen un problema: las capas intermedias (también conocidas como “ocultas” o *hidden layers*) no permiten saber exactamente cómo se produce el proceso de generalización. Por ejemplo, en nuestro caso de selección de lunares cancerígenos frente a benignos, es posible que la red acierte, pero no se sabe bien cómo realiza su elección, ni en qué tipo de parámetros se basa. En este ejemplo concreto, no parece un problema muy grave, pero imaginemos una red cuyo fin sea decidir sobre la concesión o no de préstamos bancarios: tras ser entrenada con multitud de casos basados en una gran cantidad de parámetros distintos, aprende a distinguir entre personas que han sido capaces de devolver un préstamo bancario y aquellas que no lo han sido. Una vez el programa ha sido entrenado para distinguir entre estos casos reales, puede usarse con nuevos casos para predecir si alguien tiene una mayor probabilidad de devolver o no devolver un préstamo; en caso

negativo, el banco puede denegar entonces la solicitud del cliente. Pero si le preguntan al banco por las razones del rechazo, no serán capaces de explicitarlas: la información específica que ha usado la red para tomar la decisión está distribuida en los pesos de conexiones de la capa oculta y no hay manera de saber cuáles han sido las razones concretas. Lo mismo sucede con un programa que diga a qué prisioneros se debe conceder libertad condicional o no según su experiencia con casos previos. Si el programa la aprueba o la deniega no hay manera de saber por qué lo ha hecho. Además de los problemas éticos, de este tipo y otros (se ha demostrado que los programas reproducen los prejuicios y sesgos implícitos en los datos con los que han sido entrenados, por ejemplo), esto obviamente va a ser uno de los obstáculos para utilizar el tratamiento de datos masivos con redes neuronales como herramienta para averiguar el funcionamiento de sistemas complejos. En cualquier caso, y como acabamos de decir, el uso de los *big data* es un tema altamente controvertido, por otras muchas razones adicionales. Ya hemos mencionado el espinoso problema de la privacidad. En 2016 surgió una gran controversia al descubrirse que agencias del gobierno de Estados Unidos (la NSA o *National Security Agency* o el FBI) habían accedido a millones de correos electrónicos de compañías como Yahoo, por ejemplo, para escanear su contenido en busca de material comprometido, con el objeto de combatir el terrorismo. Resulta ilusorio pensar que este tipo de prácticas han cesado; claramente, el poder de los *big data* puede ser utilizado para asuntos de gran utilidad, como este que mencionamos sobre actividades terroristas o peligrosas. Lo mismo podríamos decir de información relacionada con nuestra salud, que puede servir para diseñar programas sanitarios destinados a atajar determinados problemas. Sin embargo, a muchas personas les resulta inaceptable la utilización de datos personales por parte de agencias que no siempre exhiben el grado de transparencia adecuado. El gobierno de China recopila todo tipo de información sobre sus ciudadanos y ha llegado a establecer un “sistema de crédito social”, que consiste en la monitorización continua del comportamiento de sus ciudadanos, puntuándolo y estableciendo un sistema de “recompensa” o “castigo”. Se llega a extremos que resultan muy llamativos en las democracias occidentales: por ejemplo, las cámaras de reconocimiento facial pueden detectar a personas que crucen una calle con el semáforo en rojo, que pueden ser introducidas en una “lista negra” si son reconocidas en cinco ocasiones distintas en esta conducta incívica. Por supuesto, esto es un ejemplo extremo: es mucho más sencillo controlar violaciones de pago de impuestos, búsquedas consideradas inadecuadas en servidores, o comentarios en redes sociales, apoyo a causas críticas con el gobierno y un larguísimo etcétera. Estas recompensas o castigos pueden tener consecuencias como la facilidad o dificultad para conseguir un billete de avión o alojamiento en hoteles de prestigio, o el ingreso en un colegio u otro de tus hijos. Muchos advierten de que esta situación es muy similar al Gran Hermano de Orwell; toda esta información recopilada por el sistema está basada en estos datos masivos recopilados de manera digital. En Occidente no existe un programa explícito de manejo de los datos de esta manera, pero obviamente, la privacidad personal ha disminuido e incluso, según algunos, esta es ya un concepto ilusorio, perteneciente a un tiempo pasado.

3. El *big data* y los estudios lingüísticos

En principio, se tendería a pensar que esta revolución y las oportunidades que ofrece habrían sido bienvenidas sin ningún tipo de reticencia por parte de los lingüistas. No es

el caso, sin embargo; tal y como hemos mencionado, las dos características más relevantes de los *big data* (es decir, su volumen y su variedad) ofrecen por igual tanto oportunidades como desafíos. Examinemos ambas características por turnos.

3.1. El volumen de datos

Para empezar, en lo referente al volumen, algunos lingüistas han expresado un cierto escepticismo frente a estos estudios de grandes datos. Sus reservas vienen de varias fuentes. Por un lado, se argumenta que los lingüistas llevan ya un tiempo trabajando con *big data*, al menos, en lo referente al volumen de información. Mientras que en los años 1960-1970, corpus como el Brown o el LOB tenían un millón de palabras, en los años 80, el corpus de Birmingham/Cobuild contaba ya con 20 millones de palabras; así mismo, en los años 90, el British National Corpus contaba con 100 millones de palabras, lo que durante mucho tiempo fue considerado un estándar. Ya en el siglo XXI, The Bank of English cuenta con 645 millones de palabras y la familia de corpus TenTen ofrece tamaños mucho mayores aún: el TenTen inglés, por ejemplo, cuenta con 19.000 millones de palabras. Existen, además, otro tipo de críticas relacionadas con el volumen de datos. Esto es así porque en realidad no existe un acuerdo generalizado sobre la importancia que tiene el tamaño de un corpus. El asunto de cómo de grande debe ser un corpus es un tema muy controvertido. Existen opiniones absolutamente encontradas, como la de Krishnamurthy (2001) quien expresa claramente su opinión en el título de su artículo (“Size matters”), o la de Geoffrey Leech (1991: 10), quien opina que el tamaño no es tan importante (“Size is not all-important”). La clave de este desacuerdo se basa en la función que debe desempeñar un corpus o dicho de otro modo, la respuesta a la pregunta de hasta qué punto es posible generalizar, a partir de los datos encontrados en un corpus dado, acerca del funcionamiento general de la lengua. John Sinclair, una de las figuras fundacionales de los estudios de corpus lingüísticos (véase García-Miguel 2022), advertía de que era preciso distinguir entre “archivos” o “bases de datos lingüísticos” y “corpus” (Sinclair 1991). Una lista de palabras, un archivo de texto cualquiera, una recopilación de citas, por ejemplo, no formarían un corpus. La información textual de internet, según este autor, tampoco conformaría un corpus. Por ejemplo, los datos de frecuencia de uso en internet pueden ser muy engañosos: si buscamos en Google la frase *A long time ago in a galaxy far far away*, nos devuelve más de 800.000 resultados. Esta alta frecuencia no es representativa del uso de esta frase en el lenguaje en general, sino que se debe más bien a su inclusión en multitud de páginas que hacen referencia a su uso en la franquicia de la Guerra de las Galaxias. De una opinión parecida es Andrew Hardie, otra figura señera en la lingüística de corpus (autor de uno de los programas de procesamiento de corpus más completos y populares, CQPWeb), que, de hecho, se muestra muy crítico de entrada con los acercamientos de *big data*. Según este autor, algunos estudios basados en *big data* son apenas “pseudoinvestigación”; llega a decir que “the prospect of a *big data* revolution in linguistics is fundamentally illusory” [‘la posibilidad de una revolución de *big data* en la lingüística es fundamentalmente ilusoria’] (Hardie 2010: 23). De manera que la razón por la que el tamaño de un corpus puede ser un factor secundario es esta: existe un acuerdo generalizado en que un corpus no es una mera colección de textos, sino que se construye deliberada y cuidadosamente para que pueda servir como ejemplo representativo de una lengua, y para que los resultados extraídos de sus textos puedan servir de ejemplo de cómo funciona la lengua en general, es decir, puedan “generalizarse” y no tomarse como una característica especial del corpus del que se han obtenido los datos. Es decir, un

corpus debe ser “representativo” y “equilibrado”, con una composición cuidadosamente escogida. El problema añadido es que estos dos conceptos no han alcanzado una definición aceptada por todos y son todavía muy controvertidos. Según Leech (1991), la suposición de representatividad debe ser más bien aceptada como un “acto de fe”; autores como Tognini-Bonelli (2001) también han expresado su escepticismo, ya que no tenemos de momento manera de asegurarla o de evaluarla de manera objetiva. Discusiones de este tema pueden consultarse en Biber (1993) o Atkins et al. (1992).

En realidad, parece que el tamaño ideal de un corpus depende, en primer lugar, de su propósito, es decir, de cuál sea la pregunta de investigación que se desea responder. No es lo mismo estudiar aspectos muy generales del lenguaje que un fenómeno muy poco frecuente, que, posiblemente, no aparezca o no aparezca (o, al menos, no con la frecuencia deseada) en un corpus de tamaño medio.

En cualquier caso, y una vez presentadas las objeciones, hay que decir que también existen muchos argumentos a favor de corpus lo más grandes posibles. Un mayor tamaño facilita de manera obvia el estudio de fenómenos lingüísticos más raros, menos frecuentes, que no ocurren en corpus pequeños con la frecuencia necesaria. Igualmente, contar con un número suficiente de datos permite el estudio de un fenómeno teniendo en cuenta distintas variables; al usar distintas variables, el número de casos para el estudio debe aumentar. Por poner un ejemplo muy sencillo: si pensamos que hombres y mujeres exhiben una diferencia en la expresión de un fenómeno lingüístico, debemos recopilar un número de casos que será dividido en dos grupos según la variable “sexo”, lo que requiere el doble de casos que si no hiciéramos esa distinción. Si aumentamos el número de variables (que es potencialmente muy amplio: registro, contexto lingüístico, procedencia geográfica, frecuencia léxica, número de vecinos semánticos, valencia, y un larguísimo etcétera), se empieza a comprender por qué puede ser necesario contar con un número de ejemplos lo más grande posible.

3.2. La variedad de los datos

La otra característica de los *big data* que hemos destacado como especialmente relevante para los estudios lingüísticos es su gran variedad. Ya hemos mencionado que existen muchos tipos de datos en el *big data* y también que existe un número potencialmente muy grande (podríamos incluso decir que un número indefinidamente grande) de variables que pueden ser relevantes en el estudio lingüístico. Por ejemplo, con mucha frecuencia, los datos lingüísticos (esto es, las palabras) se anotan con información morfológica (según el tipo de morfema, por ejemplo), sintáctica (si una palabra es nombre, verbo o adjetivo) o gramatical (anotándose la función de una palabra, si es sujeto, objeto o modificador verbal). Potencialmente, hay muchas más capas de anotaciones posibles: anotaciones prosódicas, gestuales, discursivas, pragmáticas, geográficas (usando el GPS, por ejemplo), entre otras. Pensemos por un momento en la investigación sociolingüística: incluir información del hablante (su género, su edad) o del momento en que se dijo algo o el lugar geográfico, por ejemplo, son todos aspectos absolutamente importantes para una caracterización sociolingüística de las expresiones. Muy relacionado con este tema encontramos las distintas “capas” de información que se distinguen en los corpus lingüísticos. Los datos asociados con un corpus lingüístico se pueden dividir en tres tipos.

En primer lugar, tenemos los datos primarios, que son el material lingüístico en sí, es decir, las expresiones lingüísticas escritas o dichas por los hablantes. En segundo lugar,

están los metadatos. En esta segunda capa se incluye información referida a los detalles de la captura o toma de los datos: dónde se llevó a cabo, en qué momento, de dónde se extrajeron los datos –periódico, entrevista de televisión, novela...–. En el último lugar se encuentra la capa de anotación. Esto son datos añadidos por el analista o por el sistema y corresponden a información morfosintáctica, semántica, gestual, prosódica, o de cualquier otro tipo que podamos imaginar.

Esta última capa es la que ofrece mayores oportunidades de expansión y enriquecimiento, ya que es absolutamente abierta. Unos datos primarios (por ejemplo, un texto) se pueden anotar con sucesivas capas de información, que permitirán, por un lado, búsquedas en el corpus mucho más refinadas y enfocadas, y por otro, la realización de estudios cuantitativos mucho más sofisticados, que exploren correlaciones completas entre los distintos tipos de información presente en las capas de anotación.

Sin embargo, es precisamente esta promesa de enriquecimiento de la información la que trae aparejada nuevos problemas, que son críticos para llegar a un aprovechamiento real de los “grandes datos” en los estudios lingüísticos. Para empezar, la adición de determinadas etiquetas o anotaciones puede (y, de hecho, suele) estar asociada a unos supuestos teóricos concretos; es difícil encontrar etiquetas que estén “libres de teoría”, lo que limita la utilidad de ese etiquetado a los investigadores que siguen determinada teoría. Pero, al margen de este problema, el verdadero cuello de botella tiene que ver con la manera en que se añaden esas etiquetas que describen los aspectos suplementarios. La adición de etiquetas puede ser un proceso manual, automático o semiautomático. El proceso manual es el que permite un ajuste más preciso y razonado de la etiqueta que se añade; determinados tipos de anotación pueden necesitar de manera más acusada el juicio humano, como podría ser el caso de anotaciones de tipo semántico o pragmático. Sin embargo, esto conlleva un claro coste temporal: un humano puede anotar únicamente una cantidad de información determinada y, cuando estamos hablando de corpus de miles de millones de palabras que aumentan a una gran velocidad, lo más probable es que únicamente un porcentaje muy limitado del tamaño real pueda ser anotado de esta manera. Es decir, podemos tener Volumen o Variedad, pero no ambas cosas al mismo tiempo. A este problema se añade el que la anotación humana trae también aparejada problemas relativos a la “subjetividad” del analista. Esto quiere decir que se requiere una validación adicional: los mismos datos deben ser anotados por distintos anotadores y se debe efectuar una comparación de las anotaciones para ver si se alcanza el necesario nivel de “acuerdo entre jueces”, para lo que existen herramientas estadísticas que miden el nivel de acuerdo deseado (por ejemplo, la K de Kendall).

Queda entonces claro que el verdadero desafío consiste en encontrar herramientas de anotación que sean o bien completamente automáticas, o al menos, semiautomáticas, y que permitan, por tanto, a los anotadores humanos optimizar el proceso de anotación y aumentar de manera sensible el número de casos anotados. De otra manera, aunque seamos capaces de satisfacer la primera V, de “Volumen”, por muy grande que sea el volumen de datos primarios, no tendremos la V de “Variedad” más que en un grupo reducido de casos. En las siguientes secciones vamos a examinar este problema con más detalle centrandó nuestra atención en un tipo especial de corpus que intenta incluir información lingüística de distintos tipos: los corpus multimodales.

4. Un ejemplo de variedad en los datos: los corpus multimodales

La cantidad y variedad de información que está presente en la comunicación lingüística va más allá de lo puramente verbal, es decir, del simple reconocimiento de una lista de palabras (véase la Figura 6). Un alto porcentaje de la información que utilizamos de manera usual en la comunicación se recibe de manera visual, no auditiva; es el caso de los gestos, las posturas corporales, las expresiones faciales, la mirada... Por supuesto, también hay información adicional más allá del reconocimiento léxico en la señal sonora; existen toda una serie de matices de entonación que no siempre se tienen en cuenta en los análisis lingüísticos. Pensemos en un intercambio tan sencillo como alguien que dice la palabra *hola*. Dependiendo de la entonación, de la expresión facial o de la postura corporal, podemos entenderla de muchísimas maneras distintas: por ejemplo, podemos imaginar un *hola* que signifique ‘qué sorpresa verte aquí’ (asociado con una entonación y una expresión facial determinada), o que signifique ‘¿hay alguien aquí?’ (dicho con un volumen y una entonación concreta), o ‘cómo me sorprende que digas esto’, o simplemente un saludo neutro. De todos estos parámetros, el estudio de los gestos está especialmente avanzado y se ha reconocido ya su ubicuidad: al parecer, no existe ningún lenguaje humano en el que no se gesticule al hablar. Estos gestos a veces refuerzan la información presente en el mensaje verbal (actúan de manera redundante); a veces ofrecen información nueva y lo complementan; y a veces ofrecen pistas de cómo interpretar de manera correcta un mensaje con varias interpretaciones posibles. Existe toda una nueva corriente en los estudios de lenguaje dirigida a extraer información multimodal de los eventos comunicativos e integrarla en modelos más complejos del lenguaje. Este movimiento de comunicación multimodal está en plena evolución y sus fronteras están todavía por definir. Áreas como la gestualidad (y la entonación) están más desarrolladas, pero existen muchas otras áreas, más minoritarias, que están desarrollándose de manera paralela. Entre otros ejemplos, Keevalik y Ogden (2020) ofrecen una revisión de sonidos usados en la comunicación (extraídos del inglés y el estonio) y que no forman parte del repertorio fonético clásico de una lengua.



Figura 6. Ejemplos de multimodalidad en la comunicación lingüística

Un ejemplo concreto del nuevo tratamiento con datos muy variados y no incluidos en los análisis más tradicionales es el proyecto BabyCASE, que podemos usar como ejemplo para comprender tanto la utilidad y el valor explicativo de este tipo de información, como las dificultades inherentes a su uso y tratamiento. Este proyecto contiene 764 ejemplos transcritos de comportamiento no verbal, con rasgos como movimientos de cabeza (asentimientos o negaciones), sonrisas, dirección de la mirada, inclinación del cuerpo hacia delante o hacia atrás, imitaciones, risas y diversos movimientos de las manos (Brunner y Diemer 2018). Entre otros aspectos, este proyecto ha estudiado cómo se relacionan aspectos no verbales con eventos de *code-switching*: cambios de idioma de los hablantes, que pasan a utilizar en un momento dado y puntual un código distinto en una

interacción. En la Figura 7 se puede ver así las distintas clasificaciones de tipos de risas, cómo se relacionan con eventos de *code-switching*, y cuál es la posición de esa risa en el contexto en el que se cambia de idioma.

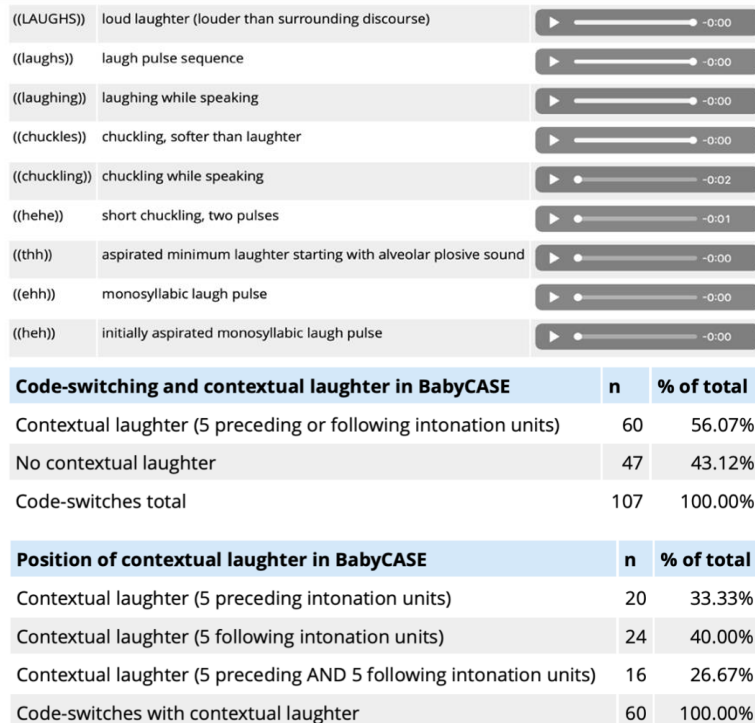


Figura 7. Ejemplos de análisis multimodal: función de distintos tipos de risas en el code-switching (Brunner y Diemer 2018)

Este estudio, por lo tanto, cumpliría una de las dos características principales del *big data* (la variedad), pero ofrece un número extremadamente limitado de ejemplos analizados. Podemos comprender, al examinar este estudio, la dificultad de extender este tipo de análisis, absolutamente dependiente de la anotación humana, a la enorme cantidad de datos disponibles. Este es, de hecho, el principal problema de los corpus multimodales. Hace ahora unos 10 años, en Knight (2011) se comentaba que apenas existían corpus multimodales cuyo tamaño fuera más allá de unos pocos miles de palabras. Y se mencionaba como ejemplo el AMI corpus con “an impressive 100 hours of video”. Estos datos sirven como referencia para la siguiente sección, en la que presentamos un corpus multimodal que puede ser tenido en cuenta como integrante real del movimiento *big data*, debido tanto a la variedad de información ofrecida como a su gran tamaño.

5. The NewsScape corpus y el Laboratorio Red Hen

Presentamos en este apartado la investigación llevada a cabo por el Laboratorio Red Hen (<https://redhenlab.org>), un consorcio internacional que agrupa a distintas universidades y grupos de investigación, y que está dedicado al estudio del lenguaje multimodal. El principal recurso del Red Hen Lab es una base de datos multimodal, conocida como la *NewsScape Library of International Television News* (a partir de ahora *NewsScape*), alojada en la biblioteca de la University of California, Los Ángeles (UCLA). Esta base

de datos consiste en la recopilación de un gran número de horas de programas de televisión, más de 500.000. Estos programas se graban junto con los subtítulos de lo que se dice en cada momento. De manera crucial, estos subtítulos se alinean de manera adicional (mediante un procedimiento conocido como *forced-alignment*): un súper ordenador de altas prestaciones es el responsable de incluir una marca temporal (cada segundo) en los subtítulos. Esto permite que el material textual que describe lo que se está diciendo pueda ser tratado como cualquier otro corpus textual, con las mismas herramientas de búsqueda. En este caso, sin embargo, los resultados de una búsqueda textual incorporan esta marca temporal, de manera que se puede buscar una palabra, una frase o una expresión gramatical compleja y los resultados incluyen un enlace al momento exacto en que se dijo dicha expresión en los archivos de vídeo. Las implicaciones de esto para el estudio multimodal son enormes. Podemos no solo saber cuántas veces, cuándo y quién ha dicho una determinada expresión, sino examinar a los hablantes en el momento de decirla, teniendo por lo tanto a nuestra disposición información multimodal como rasgos prosódicos, gestuales o de cualquier otro tipo. El corpus textual formado por esos subtítulos asociados a las 500.000 horas (aproximadas, puesto que el corpus va creciendo a un ritmo aproximado de 150 horas por día) alcanza un tamaño de unos 4.000 millones de palabras, lo que permite clasificar el estudio basado en *NewsScape* como “de big data”.



Figura 8. The Red Hen Lab (<https://www.redhenlab.org>)

5.1. Formato de los datos en Red Hen: metadatos y anotaciones

Como hemos dicho anteriormente, se ha reservado la etiqueta de “metadatos” para el etiquetado que es añadido de manera más o menos automática y que responde a información sobre los detalles de la captura de los datos en cuestión: el origen de los datos, esto es, de qué fuente se extrajeron (si era un periódico, una novela o una conversación telefónica), cuándo se realizó la toma de datos, etc. En *NewsScape* existen una serie de metadatos adaptados a su idiosincrasia. Como todos los datos se extraen de programas de televisión, los datos primarios vienen anotados de manera automática con información sobre de qué cadena de TV se extraen los datos, el momento temporal (año,

mes, día, hora, segundo). En la Tabla 1 se muestra una selección de los metadatos más básicos en NewsScape (para más información, se puede visitar la dirección www.redhenlab.org).

TOP	contiene la marca de tiempo de inicio y el nombre del archivo
COL	contiene el nombre de la colección
UID	una identificación única para la colección
PID	el episodio del programa (EP) o el ID del programa (SH) (cuando esté disponible)
AQD	el momento de la adquisición
DUR	la duración de la grabación en horas: minutos: segundos, centésimas de segundo
VID	el tamaño de la imagen del vídeo comprimido y del vídeo original
TTL	el título del evento si corresponde, o la serie, si contiene caracteres no ascii
URL	la fuente web si corresponde
TTS	el tipo de transcripción si corresponde
SRC	el lugar de grabación
CMT	un comentario agregado por la persona que programa la grabación
LAN	código de idioma ISO de tres letras (ISO 639-2T)
TTP	la página de teletexto
HED	el encabezado si está disponible, generalmente con información resumida sobre el contenido
OBT	la hora de transmisión original, cuando difiere de la hora de transmisión local “OBT Estimado” se utiliza en archivos digitalizados cuando se desconoce la hora de transmisión precisa
LBT	la hora de transmisión local, con zona horaria
END	La marca de tiempo de finalización se deriva de la hora de inicio más la duración del vídeo. Le sigue una repetición del nombre del archivo.

Tabla 1. Metadatos en NewsScape

En teoría, hemos mencionado una distinción existente entre “metadatos” (datos relativos a la recogida) y “anotaciones” (datos añadidos por anotadores humanos). Sin embargo, estas dos categorías con frecuencia se solapan, de manera que se distinguen únicamente dos capas: la primaria, de los datos en sí (texto, audio) y los metadatos. Podemos dividir los metadatos añadidos o secundarios entre los automáticos y los añadidos por anotadores humanos. En Red Hen se intenta utilizar herramientas de anotación automática a disposición de la comunidad académica; por ejemplo, programas de *software* existentes en el mercado que anotan de manera automática las “partes de la oración” de todas las palabras de un corpus. En otras ocasiones, se añaden anotaciones provenientes de equipos de anotadores humanos (véase Tabla 2).

FRM_01	marcos lingüísticos (<i>frames</i>) de la FrameNet 1.5 por medio del software Semafor 3.0-alpha4
GES_02	gestos de tiempos etiquetados manualmente por el grupo de investigación del Daedalus Lab
GES_03	gestos etiquetados manualmente con ELAN
NER_03	reconocimiento de nombres de entidades (usando el Stanford NER tagger 3)
POS_01	categorías gramaticales inglesas con dependencias, usando MBSP 1.4
POS_02	categorías gramaticales inglesas usando el Stanford POS tagger 3.4
POS_03	categorías gramaticales alemanas usando Pattern.de
POS_04	categorías gramaticales francesas usando Pattern.f
POS_05	categorías gramaticales españolas usando pattern.es
SEG	fronteras entre historias por Weixin Li, UCLA
SEG_00	fronteras de anuncios, usando información de CCEXtractor 0.74
SEG_01	detección de anuncios por Weixin Li
SEG_02	Fronteras entre historias por Rongda Zhu, UIUC
SMT_01	Análisis de sentimientos usando Pattern 2.6
SMT_02	Análisis de sentimientos usando SentiWordNet 3.0
DEU_01	Traducción automática de alemán a inglés

Tabla 2. Algunas etiquetas de metadatos añadidas en NewsScape

Uno de los principales objetivos de Red Hen es desarrollar herramientas que permitan automatizar el proceso de etiquetado manual, que, como hemos dicho, es el verdadero cuello de botella de los corpus multimodales. En la actualidad, existen diversos procedimientos de etiquetado (que van del etiquetado manual, realizado por grandes equipos de investigadores, o de etiquetados semiautomáticos, en los que se utiliza un *software* especial para acelerar o facilitar el proceso de anotación), hasta distintas propuestas para un etiquetado automático. A continuación explicamos algunos de estos procesos:

(i) Etiquetado manual

El etiquetado manual se lleva a cabo por parte de equipos dedicados a anotar un determinado comportamiento multimodal. Por ejemplo, el equipo del Daedalus Lab, de la Universidad de Murcia, ha ido recopilando información y tiene una base de datos de unos 10.000 clips anotados gestualmente. Esta base de datos está dedicada a estudiar los gestos espaciales que realizamos al articular expresiones temporales; por ejemplo, el gesto lateral realizado mientras decimos “from beginning to end” (Pagán et al. 2020; Valenzuela et al. 2020). Los clips en los que aparecen este tipo de expresiones han sido clasificados según la visibilidad de los gestos manuales, según el eje utilizado en el patrón gestual, su direccionalidad, la mano utilizada y una variedad de aspectos adicionales (véase también Alcaraz y Valenzuela 2021 para una relación entre la distancia temporal presente en la expresión lingüística –*distant future* vs *near future*– y la distancia física del gesto que acompaña a estas expresiones). La base de datos de la que hablamos ha sido

recopilada de manera manual en un periodo de unos 5-7 años. Es por ello por lo que es necesario intentar acelerar el proceso y aumentar el número de casos anotados.

(ii) Etiquetado manual ayudado por software: el Red Hen Rapid Annotator

Gracias a la colaboración con el lingüista computacional Peter Uhrig, así como por la participación del Red Hen Lab en el proyecto Google Summer of Code, se ha diseñado un *software* que optimiza el tiempo de anotación humana. Al anotador se le presentan los clips de manera automática con una serie de menús con opciones y sus elecciones pasan a grabarse de manera automática en la base de datos, acelerando de esta manera el tiempo de anotación en un factor de x10. La Figura 9 ofrece un ejemplo de este programa para anotar características de una imagen.



Figura 9. Red Hen Rapid Annotator

(iii) Aprendizaje por máquina: visión computacional

Se han realizado también experimentos en los que los clips de vídeo han sido procesados por medio de un algoritmo de visión computacional, que de nuevo optimiza de manera clara el tiempo dedicado a la anotación manual. Por ejemplo, los anotadores humanos deben descartar aquellos clips de vídeo en los que se pronuncia una expresión dada, pero no aparece ningún humano en pantalla (lo que se conoce como “voz en *off*”: una persona habla describiendo atascos de tráfico, por ejemplo, pero las imágenes no muestran a esta persona hablando, sino imágenes de tráfico con carreteras con largas colas). El programa desarrollado por Sergiy Turchyn (Turchyn et al. 2018) era capaz de detectar en un clip la presencia o ausencia en la pantalla de un hablante humano, descartando de manera automática los casos de “voz en *off*” y agilizando de manera muy clara el tiempo de anotación del equipo humano. Un estudio piloto en la Universidad de Navarra comprobó que el uso de este software reducía a la mitad (de 1:07:55 a 00:33:47 de media) el tiempo necesario para llevar a cabo la anotación manual.

(iv) Open Pose

En último lugar, en estos momentos el equipo del Red Hen Lab está explorando la utilización del *software* OpenPose (Cao et al. 2021). Este programa de visión computacional identifica puntos del cuerpo humano, especialmente las articulaciones y otros puntos movibles del cuerpo, así como los ojos, o las cejas, y les asigna un número, tal y como se ve en la Figura 10.

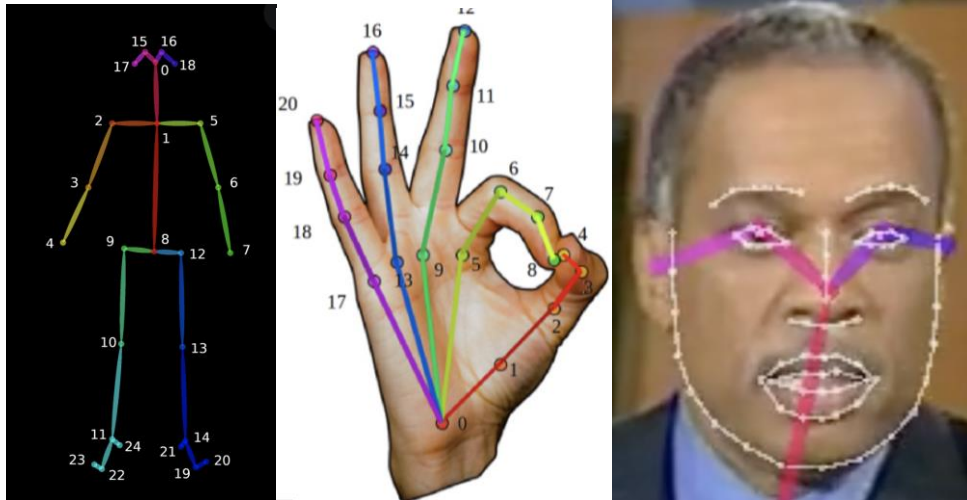


Figura 10. Algunos puntos reconocidos por el programa OpenPose

Los clips de vídeo que resultan de interés para el programa de investigación son de esta manera procesados por el programa OpenPose, que permite la visualización con la información superpuesta o simplemente con los puntos aislados (véase Figura 11).



Figura 11. Modos de visualización de la información en OpenPose

A partir de este análisis, el programa genera archivos que van indicando en qué posición espacial (es decir, cuáles son las coordenadas x e y) se encuentra el punto correspondiente a una determinada parte del cuerpo en un *frame* de un vídeo. Esto quiere decir que es posible realizar un seguimiento del desplazamiento espacial de, digamos, el punto 4, correspondiente a la muñeca izquierda, y comprobar su trayectoria a medida que transcurre el tiempo del clip. Obviamente, esto genera una ingente cantidad de datos (recordemos que un segundo de vídeo contiene unos 30 frames), que tienen que ser sometidos a un cuidadoso análisis por medio de programas que lidien con el “ruido” que

el programa inevitablemente genera (el movimiento de un punto puede verse momentáneamente interrumpido por un obstáculo visual, puede haber varias personas hablando al mismo tiempo, puede haber cambios de cámara que requieran recalculas las posiciones relativas de los puntos, y un largo etcétera). En cualquier caso y una vez superadas estas dificultades técnicas, cruzando estos datos de movimiento detectado automáticamente con los datos de audio de cuándo se empieza a decir algo, estamos en disposición de identificar automáticamente gestos, de manera que podríamos ser capaces de realizar búsquedas como “encuentra a alguien diciendo *en los meses venideros* mientras hace un gesto con la mano derecha de izquierda a derecha”, o incluso de buscar gestos de manera automática y estudiar qué tipo de expresiones lingüísticas son las que se asocian con su realización. En estos momentos, el grupo Daedalus Lab está trabajando en este problema, que permitiría dar el salto a un acercamiento verdaderamente de *big data* a los datos multimodales, puesto que la mayor parte del procesamiento de anotación gestual pasaría a automatizarse por medio del algoritmo de OpenPose. La unión de este tipo de información con un análisis prosódico de la señal sonora por medio de programas tipo Praat (Boersma y Weenik 2021) puede ser realmente la puerta que por fin permita llevar el análisis multimodal al siguiente estadio, iniciando la etapa de *big data* en lingüística en su sentido más completo.

6. Conclusiones

A lo largo de este trabajo hemos visto cómo los paralelismos del *big data* en su uso más extendido y generalizado y los *big data* en Lingüística son más bien aproximados. De hecho, sus propósitos son distintos: mientras que el *big data* en general es una empresa de tipo eminentemente práctica, un problema de ingeniería; el objetivo de los *big data* en lingüística es muy distinto: la lingüística es una disciplina académica cuyo propósito es entender los mecanismos subyacentes a la comunicación lingüística, y el uso de *big data* debe servir siempre a ese objetivo de comprensión del fenómeno. De hecho, existen áreas de la lingüística en las que el uso del *big data* en su versión ingenieril ha resultado de gran utilidad: es el caso de los sistemas de reconocimiento de habla, o los sistemas de traducción automática, que han mejorado su funcionamiento de manera notable en los últimos años gracias al entrenamiento de programas con grandes datos (aunque esto no haya llevado aparejado un incremento en la comprensión de cómo se produce el fenómeno).

Claramente, la utilidad de los *big data* para la empresa lingüística es indudable, siempre que se sea consciente de su papel, sus ventajas y sus problemas. En este sentido, como cualquier otra herramienta, su utilidad depende en gran medida del uso que se le quiera dar. Pero, al margen de estas reticencias y precauciones, resulta evidente que el uso de los *big data* puede contribuir de manera decidida a la construcción de modelos mucho más complejos y completos del lenguaje. Como viene siendo la costumbre en los últimos años, los datos provenientes de este tipo de fuente deberán entrar en la “rueda” del resto de herramientas metodológicas que conforman la ciencia cognitiva (métodos conductuales –véase Igoa (2022)–, de seguimiento ocular –*eye-tracking*, véase Álvarez García (2022)– o provenientes de la neuropsicología), contribuyendo de esta manera a afianzar la noción de “evidencia convergente” ya establecida en las ciencias cognitivas: cuando diferentes tipos de evidencia empírica recogida desde diferentes perspectivas apuntan a un mismo tipo de explicación.

7. Agradecimientos

Este trabajo ha sido realizado gracias al apoyo del Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación y fondos FEDER/UE funds (grant number PGC2018-1551 097658- B-100).

8. Referencias

- Alcaraz Carrión, Daniel; Valenzuela, Javier. 2021. Distant time, distant gesture: speech and gesture correlate to express temporal distance. *Semiotica* 241. DOI: 10.1515/sem-2019-0120
- Álvarez García, Esther. 2022. Lo que esconden tus ojos: la metodología eye-tracking aplicada al estudio del lenguaje. *Estudios de Lingüística del Español* 45: 205-239.
- Atkins, Sue; Clear, Jeremy; Ostler, Nicholas. 1992. Corpus design criteria. *Literary and Linguistic Computing* 7.1: 1-16.
- Biber, Douglas. 1993. Representativeness in Corpus Design. *Literary and Linguistic Computing* 8.4: 243-257.
- Boersma, Paul; Weenink, David. 2021. *Praat: doing phonetics by computer* [Computer program]. Version 6.1.50.
- Brunner, Marie-Louise; Diemer, Stefan. 2018. “You are struggling forwards, and you don’t know, and then you ... you do code-switching...” – Code-switching in ELF Skype conversations. *Journal of English as a Lingua Franca* 7.1: 59-88.
- Cao, Zhe; Hidalgo, Gines; Simon, Tomas; Wei, Shih-En; Sheikh, Yaser. 2021. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43 .1: 172-186.
- García-Miguel, José M. 2022. Lingüística de corpus: de los datos textuales a la teoría lingüística. *Estudios de Lingüística del Español* 45: 11-42:
- Hardie, Andrew. 2010. *Big data* in language studies: from cargo-cult science to phantom revolution. Conferencia plenaria en el 7 Congreso de AELINCO 2015, Universidad de Valladolid.
- Keevallik, Leelo; Ogden, Richard. 2020. Sounds on the Margins of Language at the Heart of Interaction. *Research on Language and Social Interaction* 53.1: 1-18. DOI: 10.1080/08351813.2020.1712961
- Knight, Dawn. 2010. The future of multimodal corpora. *Revista Brasileira de Linguística Aplicada* 11.2: 391-415.
- Krishnamurthy, Ramesh. 2001. Size Matters: creating Dictionaries from the World’s Largest Corpus. *8th Annual KOTESOL Conference Proceedings*. Taegu: KOTESOL: 169-180.
- Igoa, José Manuel. Las tareas conductuales en la investigación sobre el procesamiento del lenguaje. *Estudios de Lingüística del Español* 45: 133-158.
- Leech, Geoffrey. 1991. The state of the art in corpus linguistics. En K. Aijmer y B. Altenberg, eds. *English Corpus Linguistics*, Londres: Longman, pp. 8-29.
- Olza, Inés; Valenzuela, Javier; Pagán-Cánovas, Cristobal. 2017. Automatic visual analysis and gesture recognition: Two preliminary pilots. Universidad de Navarra: Instituto Cultura Sociedad.

- Pagán Cánovas Cristóbal; Valenzuela Javier; Alcaraz Carrión Daniel; Olza Inés; Ramscar Michael. 2020. Quantifying the speech-gesture relation with massive multimodal datasets: Informativity in time expressions. *PLOS ONE* 15.6: e0233892.
- Rumelhart, David E.; McClelland, James L.; PDP Research Group. 1986. *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1.* Cambridge: MIT Press.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work.* Amsterdam: Benjamins.
- Turchyn Sergiy; Olza Moreno, Inés; Pagán Cánovas, Cristóbal; Steen, Francis F; Turner Mark; Valenzuela, Javier; Ray, Soumya. 2018. Gesture Annotation with a Visual Search Engine for Multimodal Communication Research. En *The Thirtieth AAAI Conference on Innovative Applications of Artificial Intelligence (IAAI-18)* [Internet]. 2018. <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/viewFile/16703/16398>
- Valenzuela, Javier; Pagán-Cánovas, Cristóbal; Olza, Inés; Alcaraz, Daniel. 2020. Gesturing in the wild: spontaneous gestures co-occurring with temporal demarcative expressions provide evidence for a flexible mental timeline. *Review of Cognitive Linguistics* 18.2: 289-316. DOI: 10.1075/rcl.00061.val

Entrevistas mediadas en sociolingüística cualitativa. La aplicación en la etnografía en línea

Nadège Juan

Université de Perpignan Via Domitia - Université de Franche-Comté

nadege.juan@univ-fcomte.fr

Yvette Bürki

Universität Bern

yvette.buerki@rom.unibe.ch

Resumen

Este artículo se centra en la recolección de datos sociolingüísticos a través de entrevistas mediadas, es decir, aquellas que utilizan artefactos de las Nuevas Tecnologías de la Información y la Comunicación (NTICs), como el Smartphone y sus servicios de mensajería instantánea. Se propone una reflexión crítica sobre la noción cara a cara que ha definido tradicionalmente la entrevista, al tiempo que se exponen los aspectos positivos y desafíos metodológicos que conlleva la entrevista mediada. El estudio se divide en dos grandes partes. En la primera, de carácter más general, se explica la manera como se entiende la entrevista como herramienta heurística en estudios de tipo cualitativo y se la sitúa histórica y metodológicamente en el marco de la sociolingüística. En la segunda parte, se expone y se ejemplifica de manera práctica cómo se puede trabajar con la entrevista mediada en estudios sociolingüísticos cualitativos de carácter etnográfico y en línea. En esta vertiente de investigación sociolingüística, la entrevista se utiliza ante todo para analizar actitudes e ideologías lingüísticas. Los resultados de este estudio demuestran que la mediación tecnológica no supuso ningún impedimento comunicativo y que permitió recoger datos que enriquecieron el estudio. Pero al mismo tiempo se constata que para que las entrevistas mediadas funcionen y sean productivas, es necesario realizar un trabajo previo entre las personas participantes e investigadoras para promover un sentimiento de confianza. Se recomienda por ello invertir el tiempo necesario en este paso, que es decisivo para poder acceder a las entrevistas en línea. Asimismo, el estudio ha mostrado la importancia de la flexibilidad en el trabajo de campo, el cual requiere adaptarse a los requerimientos de las y los participantes, siempre y cuando estos no mellen los objetivos fundamentales de la investigación. A nivel ético, este estudio llama la atención sobre el hecho de que los datos en línea no están a libre disposición de los investigadores: declarar su presencia y trabajar con el acuerdo de las personas involucradas siempre y cuando se pueda es fundamental. Cada paso metodológico debe planificarse pensando en la necesidad de proteger a las personas que participan y de respetar su voluntad.

Palabras clave: entrevista mediada; sociolingüística cualitativa; etnografía; digital.

Abstract

This article focuses on the collection of sociolinguistic data through mediated interviews, i.e. those using New Information and Communication Technologies (NICTs) artefacts, such as the Smartphone and its instant messaging services. It proposes a critical reflection on the face-to-face notion that has traditionally defined the interview, while exposing the

positive aspects and methodological challenges of the mediated interview. The study is divided into two main parts. The first, more general part explains how the interview is understood as a heuristic tool in qualitative studies and situates it historically and methodologically within the framework of sociolinguistics. In the second part, it is explained and exemplified in a practical way how the mediated interview can be used in online ethnographic sociolinguistic studies. In this afore-mentioned strand of sociolinguistic research, the interview is primarily used to examine linguistic attitudes and ideologies. The results of this study show that technological mediation was not a communicative constraint and that it made it possible to collect the expected data. At the same time, however, it is clear that in order for mediated interviews to work and be productive, it is necessary to carry out preliminary work between participants and researchers in order to promote a feeling of trust. It is therefore recommended to invest the necessary time in the online presentation by the researcher. Furthermore, this study has shown the importance of flexibility in fieldwork, which needs to be adapted to the requirements of the participants, as long as these do not undermine the fundamental objectives of the research. On an ethical level, this study draws attention to the fact that online data are not freely available to researchers: declaring their presence and working with the agreement of the people involved whenever and wherever possible is essential. Each methodological step must be planned with the need to protect the people involved and to respect their wishes.

Keywords: mediated interview; qualitative sociolinguistic; ethnography; digital.

1. Introducción

La entrevista no es, por supuesto, la única manera de acceder a datos para realizar estudios sociolingüísticos de diversa índole. No obstante, es uno de los medios más empleados desde que Labov la popularizó en 1966, hasta tal punto que muchos y muchas sociolingüistas utilizan el término *entrevista sociolingüística* para denominar a la prototípica en el ámbito variacionista (véase apartado 2). Por ello, vale la pena recordar que hoy en día la sociolingüística constituye una disciplina lingüística vastísima que alberga no solo la vertiente inicial variacionista, sino otras, cercanas más bien a posturas epistemológicas etnográficas y antropológicas que emplean la entrevista como herramienta heurística para acceder sobre todo a datos metalingüísticos y metadiscursivos¹.

La entrevista se ha caracterizado en sociolingüística por ser un método de recolección de datos en una interacción cara a cara grabada (Feagin 2013: 26). Con todo, la sociolingüística también ha empleado desde sus inicios interacciones mediadas para recabar datos. El advenimiento de la pandemia por la COVID-19 supuso un problema para investigadores e investigadoras que estaban utilizando la entrevista como método de recolección de datos y, por eso, (re)abrió el debate académico sobre la posibilidad de emplear entrevistas mediadas (Marcón 2021).

En este artículo se analiza la recopilación de datos sociolingüísticos a través de entrevistas mediadas que utilizan artefactos de las Nuevas Tecnologías de la Información y la Comunicación (NTICs) como por ejemplo el Smartphone y sus servicios de mensajería instantánea.

Se propone una reflexión crítica sobre la noción *cara a cara*, que ha definido tradicionalmente la entrevista, al tiempo que se tematizan aspectos positivos y desafíos metodológicos que conlleva la entrevista mediada. El artículo se ha estructurado de la siguiente manera: después de explicar de qué manera se entiende aquí la entrevista ontológicamente y como herramienta heurística, se realiza un recorrido histórico sobre el empleo de la entrevista mediada en sociolingüística, para exponer luego los usos novedosos como técnica de recolección de datos cualitativos y cuantitativos según nuevos estudios realizados en las ciencias sociales y en sociolingüística en particular². A continuación, se presenta y se explica un estudio de caso que emplea la entrevista mediada en la sociolingüística etnográfica en línea. Se describe este estudio, sus objetivos, su metodología y pautas éticas en relación con las entrevistas en este campo. Seguidamente se explica cómo y por qué se usaron entrevistas mediadas en el estudio, y se muestra mediante ejemplos concretos, que permitieron obtener resultados interesantes. En la última parte, se formulan unas recomendaciones destinadas a los investigadores y las investigadoras que deseen incluir entrevistas mediadas en su estudio etnográfico en línea.

2. La entrevista (mediada): ontología y método

Si bien el tema de este capítulo es la entrevista mediada, es importante reflexionar, aunque sea de forma breve, cómo se entiende esta herramienta metodológica partiendo de la perspectiva de la sociolingüística cualitativa.

La entrevista prototípica se define como un evento comunicativo cara a cara en el que participan dos o más personas para hablar sobre un tema específico. La interacción es por lo general de tipo diádico, con roles claramente predeterminados: el del entrevistador (que es quien pregunta) y el entrevistado (quien responde) (Gibson y Hua 2016: 182). Tiene diferentes formatos (estructurada, semi-estructurada o libre) y constituye uno de los métodos más populares para recolectar datos en sociolingüística (Davis 2013).

La entrevista puede servir como fuente de datos conversacionales reales para realizar estudios estructurales, variacionistas e interaccionales. Dentro de esta categoría se encuentra la así denominada entrevista sociolingüística, diseñada por Labov (1966) para la elicitación de cinco estilos lingüísticos diferentes: informal, cuidado, lectura, lista de palabras y pares mínimos (Labov 1972 y Becker 2013: 91-100). La entrevista puede servir también para adquirir datos centrados en el contenido, sobre todo, en estudios sociolingüísticos de tipo cualitativo, como es el caso que aquí nos ocupa. En estos casos, el método se emplea para obtener más información, para explicitarla, y/o para explorar aspectos de la investigación a los que solo se puede acceder indirectamente en la interacción, como, por ejemplo, las actitudes lingüísticas e ideologías expresadas por los entrevistados (Codó 2008: 164). Estas posturas son esenciales para los procesos de interpretación que se realizan en la investigación, pues sirven para validarlas o cambiarlas (Codó 2008: 162). Por eso, la entrevista goza de enorme popularidad como herramienta metodológica de estudios cualitativos desde las primeras décadas del siglo XX (Fontana y Frei 1994).

A pesar de su amplio uso para recolectar datos, la entrevista también se ha visto como un “problema”. Por un lado, están quienes tratan de borrar el contexto interaccional de la entrevista, creyendo que es posible hacer que los participantes se olviden del evento para que los entrevistadores puedan acceder a su comportamiento “natural”. En el lado opuesto

están quienes sostienen que las entrevistas son contextos “no auténticos” y “artificiales” para la recogida de datos y que, por tanto, es mejor evitarlas por completo (véase De Fina y Perrino 2011). En relación con esta última postura está también la cuestión de la “verdad”, la cual pone en duda si realmente las personas entrevistadas exponen de forma sincera sus creencias e ideologías o no.

En este trabajo se siguen las propuestas de Márquez Reiter (2018: 55; también se recomienda consultar Bürki 2020: 4) en que las entrevistas constituyen encuentros sociales situados y una práctica social en sí misma. Se entienden, por tanto, como lugares dinámicos en los que las ideologías, las creencias normativas conscientes o inconscientes sobre un individuo, un grupo o una sociedad pueden formarse, reconfigurarse, reforzarse, y los discursos hegemónicos pueden impugnarse o respaldarse. De acuerdo con los planteamientos posestructuralistas y construccionistas, nosotras creemos que existe una verdad *per se*, sino posturas que van emergiendo y se van construyendo a lo largo de la interacción (Canut et al. 2018: 115). Por esta razón, no resulta plausible que las investigadoras se mantengan “neutras” en aras de la “objetividad” sin que justamente se termine violando la dinámica de una interacción natural (Briggs 1986: 39). Pensar de antemano que se debe abordar la entrevista de manera aséptica para salvaguardar esta supuesta objetividad acarrea el riesgo de generar una relación social en la que la investigadora se imponga como “observadora” desde arriba a la “informante” y en la que estas personas, más que como interactantes, se entiendan como fuentes de producción de datos (De Fina y Perrino 2011: 4). Así solo se contribuirá a hacer mayor la distancia entre entrevistadora y entrevistada.

Vale la pena recordar que la entrevista es un tipo de práctica social que nunca es totalmente espontánea, sino que obedece a un propósito concreto, determinado de antemano por la investigadora, y que tiene lugar entre ella y otra persona que por lo general no conoce. Para utilizar la entrevista respetando su calidad de encuentro interaccional real es importante entender que las propias investigadoras son parte de la producción de datos. Como señalan Canut et al. (2018: 116), queramos o no, las interlocutoras nos posicionan constantemente, nos asignan lugares, nos categorizan y se expresan o actúan en función de esos lugares asignados, los cuales, a su vez, pueden ir modificándose a lo largo de la interacción. Más que las oposiciones de objetividad versus subjetividad o de verdad versus mentira es fundamental el acto de reflexividad constante que deben realizar las investigadoras sobre el valor de las palabras y la interpretación que estas les dan en el marco de la investigación (De Fina y Perrino 2011: 3).

Para resumir, en este artículo se define la entrevista como un evento social interactivo real en la que toman parte dos personas que conversan sobre temas o aspectos que interesan a las investigadoras y así lo hacen saber a sus interlocutores. Con esta manera de entender la práctica social de la entrevista se busca establecer una relación personal más simétrica que nos permita mantener también el vínculo para continuar con nuestra investigación etnográfica en el futuro.

Como el tema aquí tratado es la entrevista mediada, cabe llamar la atención sobre un aspecto fundamental: la distancia física, que suele mencionarse como rasgo distintivo de este tipo de entrevista frente a la entrevista cara a cara. Pero como se ha señalado anteriormente, la entrevista, en tanto que práctica discursiva, ya de por sí implica una distancia inicial entre los participantes. En el caso de la entrevista como género, la distancia que implica conversar con una persona que no se conoce (bien) a través de roles

predeterminados de antemano (entrevistador y entrevistado) es, en primer lugar, intersubjetiva y, por lo tanto, psicológica, que le tocará atravesar al entrevistador y que no se rompe por el simple hecho de realizarse una entrevista cara a cara. Dicho de otro modo, la distancia física no constituye *per se* un obstáculo mayor que otras distancias. Como reconoce Marcón (2021: 90-91), en la entrevista mediada varían las características mediales y modales con respecto a las que se dan cara a cara, pero esto no quiere decir que la distancia creada por la tecnología sea un muro que actúe como barrera comunicativa infranqueable. Es más, como constata Ellingson (2012: 529), es fundamental comprender que, aunque las entrevistas puedan parecer descorporeizadas por la tecnología, los cuerpos de las investigadoras y las participantes no desaparecen ni se vuelven irrelevantes, sino que siguen estando presentes e implicados de formas complejas, pues de acuerdo con el medio se infieren señales corporales, y las emociones se manifiestan mediante diferentes modos. Aquí, como en la entrevista en presencia, entran a funcionar la reflexividad, la colaboración y, ante todo, la capacidad de la investigadora de dejarse llevar por la dinámica de la conversación para que el evento comunicativo se considere logrado.

3. La entrevista mediada. Una perspectiva diacrónica

Como es bien conocido, William Labov fue el impulsor de la entrevista como método para recabar datos en el campo de la sociolingüística variacionista. Además de las entrevistas cara a cara, Labov también utilizó las entrevistas mediadas por teléfono para obtener información fonética a través de conversaciones reales. El sociolingüista incluyó este tipo de entrevistas en su estudio sobre el Lower East Side (1966; Feagin 2013: 26) cuando los hablantes rechazaron ser entrevistados cara a cara o cuando, debido a problemas logísticos, no pudieron ser entrevistados en persona. Años más tarde, como parte de su estudio sobre el habla de Filadelfia, Labov también incorporó un muestreo aleatorio de entrevistas por teléfono para complementar el estudio llevado a cabo en el vecindario. Como explica Feagin (2013: 35), Labov aprovechó las entrevistas por teléfono para poder cubrir amplios territorios en relativamente poco tiempo. En esta línea, otra sociolingüista, Ruth Herold (1990), condujo entrevistas por teléfono en un estudio sobre la variación lingüística en Pensilvania. Herold explica detalladamente de qué manera seleccionó las localidades y los individuos buscando en el listín telefónico. Con estas herramientas pudo reunir una muestra representativa con datos de personas locales que cumplieran con el perfil de hablantes que había diseñado para su investigación. Otro estudio de gran escala que utilizó la entrevista por teléfono fue aquel sobre los efectos de la modernización en el cambio lingüístico del inglés de Tejas (Bailey y Bernstein 1989; Bailey y Dyer 1992; Bailey et al. 1997). En Bailey y Dyer (1992), por ejemplo, se empleó la entrevista telefónica para crear una muestra aleatoria de datos de forma eficiente y relativamente económica, pues utilizaron la encuesta cuatrimestral del estado de Tejas de septiembre de 1989 a la que acoplaron preguntas (el así denominado *piggy-backing method*) de tipo sociolingüístico relacionadas con el empleo léxico. A través de este método se accedió a un corpus de aproximadamente 500 horas de habla que supuso una base de datos con enorme información demográfica y que se utilizó para análisis de tipo sintáctico y discursivo (Bailey y Dyer 1992: 4).

En 1999 Labov volvió a emplear en el proyecto Telsur (*Telephone Survey of North American English*)³ la entrevista telefónica para crear la base de datos que nutre el *Atlas*

of North American English. Las entrevistas estuvieron diseñadas para obtener la aparición de las variables fonéticas utilizadas en conversaciones espontáneas en los Estados Unidos y en la parte anglohablante de Canadá. La muestra consta de casi 700 hablantes, seleccionados para representar 145 áreas urbanas con una población superior a 200.000 habitantes. Metodológicamente, este procedimiento permitió cubrir una región amplísima, por lo que se pudo ofrecer una panorámica general en el ámbito fonológico de esta región.

En sociolingüística hispánica, por su parte, casi no se han encontrado estudios que documenten entrevistas mediadas por teléfono⁴. Una excepción es el trabajo de Rigatuso (2019) en el campo de la pragmática variacional, quien señala en la parte metodológica que, para recopilar el corpus,

[s]e trabajó con interacción comercial cara a cara y telefónica, registrándose intercambios cuya duración oscila entre un minuto y más de media hora, según factores situacionales y prácticas comerciales (Rigatuso 2019: 111).

Probablemente la falta de documentación en sociolingüística hispánica con respecto a esta técnica no se deba a que no haya sido empleada, sino más bien a que en la tradición hispánica no se ha solido desarrollar la parte metodológica con exhaustividad, en la que se cuente con detalle los pasos realizados, la forma de recolección de datos, los problemas y las dificultades del diseño metodológico y de su aplicación. Felizmente, con iniciativas como la de este monográfico, cada vez hay más conciencia de la necesidad de documentar y explicar a fondo el diseño metodológico empleado en la investigación.

Un fenómeno que ha contribuido notablemente al desarrollo de nuevas metodologías en sociolingüística es Internet, con los avances tecnológicos que ello ha conllevado. A lo largo del siglo XXI, la difusión de Internet a nivel global ha ido avanzando a pasos agigantados. Para finales de 2021, Internet llegará a un 67.7 % (5098 millones de personas) de la población mundial⁵. A medida que Internet se ha ido extendiendo, ha pasado a utilizarse como campo, como herramienta y como medio de investigación en las ciencias sociales, abriendo vías metodológicas innovadoras. Pero, sobre todo, ha alterado la naturaleza del contexto en el que puede tener lugar la investigación, que se ha extendido al espacio virtual. Esto implica reflexionar también sobre cómo se construye el conocimiento en y a través de lo digital, lo cual incluye los diseños metodológicos de investigación como la entrevista, que cambian y se redimensionan. También las revoluciones tecnológicas han marcado el ritmo y el desarrollo de las técnicas de recogidas de datos, hecho que se aplica tanto a las herramientas utilizadas (desde el cuestionario inicial a los magnetófonos posteriores, hasta las grabadoras digitales y portátiles y el micrófono inalámbrico ultrasensible de la actualidad) como al canal de comunicación (Hernández Campoy y Almeida 2005: 140).

En cuanto al canal, en Marcón (2020) se identifica, además del cara a cara y el telefónico, los de naturaleza (audio)visual que ofrecen nuevas formas para establecer la comunicación con las entrevistadas y los entrevistados. Dependiendo del canal, Internet permite establecer contacto de manera asincrónica⁶ (email, servicios de mensajería) y sincrónica (servicios de telefonía, videoconferencias y chats). Además de plataformas para videollamadas como Skype, FaceTime y Zoom, son altamente populares los servicios de mensajería electrónica como WhatsApp. Asimismo, las redes sociales como Facebook, Instagram y Twitter ponen a disposición de sus usuarios servicios de

mensajería que pueden servir tanto para establecer el contacto con posibles entrevistados y entrevistadas como para realizar entrevistas. Todas estas plataformas privadas pueden utilizarse en diversos dispositivos como ordenadores, tabletas y teléfonos móviles, que, por lo general, vienen equipados con excelente calidad tecnológica (pantallas, altavoces, micrófonos de alta calidad, cámaras y diversos sensores) y son multimodales, es decir permiten emplear de manera sincrónica varios modos durante la comunicación como el sonido, el vídeo y la escritura. Estas características permiten replicar en gran medida la situación de la interacción cara a cara, como, por ejemplo, la habilidad de transmitir y responder tanto de manera verbal como no verbal (Lo Iacono et al. 2016; Archibald et al. 2019). Sobre todo, fue durante el aislamiento social impuesto por la pandemia de COVID-19 cuando más se explotaron las plataformas de videoconferencia, porque permiten entrevistar a personas evitando la proximidad física, pero contemplando una relación sincrónica y multimodal.

De acuerdo con la incipiente bibliografía existente, las entrevistas a partir de plataformas audiovisuales muestran ventajas, pero también retos. Entre los aspectos positivos destacan, como ya ocurría con las entrevistas por teléfono del siglo pasado, el bajo coste, sobre todo si es necesario entrevistar a personas que viven en zonas alejadas o apartadas o es necesario cubrir extensas zonas geográficas. Esto hace de la entrevista a través de estas plataformas una herramienta idónea para proyectos con poco presupuesto (Archibald et al. 2019; Leemann et al. 2020). Desde el punto de vista logístico, también resultan ser muy eficientes, puesto que no hay que buscar el lugar apropiado para realizar la entrevista, particularmente si se trata de entrevistas que recogen datos fonéticos, los cuales son muy sensibles a las grabaciones (Leemann et al. 2020). La mayor flexibilidad que ofrecen las entrevistas mediadas se desvela como un factor decisivo para que las personas accedan a participar. Este es un aspecto que ya mencionaba Labov (1966) y que se vuelve a poner de relieve en los estudios actuales (Archibald et al. 2019; Leemann et al. 2020). Sin embargo, no puede desestimarse el hecho de que no toda la población tiene aún acceso a Internet, de modo que este método puede ser difícil de aplicar si las entrevistas están previstas entre personas de otras latitudes (por ejemplo, Latinoamérica o África) con menos recursos económicos (James y Busher 2012).

En cuanto a los desafíos metodológicos, se mencionan aspectos interaccionales como la capacidad de establecer de forma mediada empatía entre entrevistadoras y personas entrevistadas, un aspecto en el que no parece haber acuerdo entre los investigadores (Cater 2011). A este desafío se unen los problemas para leer señales no verbales como consecuencia de una conexión retardada (Weller 2015). Asimismo, otro de los retos reside en los problemas de conexión generados en Internet, entre los que destacan la pérdida de conexión y el congelamiento de la imagen (Archibald et al. 2019; Leemann et al. 2020), que suscitan frustración entre los participantes, particularmente si duran varios minutos. También se debe tomar en cuenta la capacidad de grabar directamente las sesiones, que ofrece Zoom, pero no otros programas como FaceTime o Skype (Archibald et al. 2019) y que obligan por tanto a buscar otras formas alternativas de archivo que pueden mermar la calidad de la grabación.

Otro factor negativo importante que debe tenerse presente es el grado de alfabetización electrónica de los participantes (James y Busher 2012; Marcón 2021), principalmente si se trata de coleccionar muestras estratificadas por edad y nivel educativo, o si se contemplan diferentes tareas que obligan a utilizar otro dispositivo adicional como una aplicación electrónica (Leemann et al. 2020). Finalmente, en el caso de entrevistas que

coleccionan datos fonéticos, la calidad de la grabación es un aspecto fundamental. Por eso debe comprobarse que efectivamente el medio utilizado cuente con una tecnología que asegure una excelente calidad fonética y acústica para la reproducción posterior de los datos grabados en línea (Leemann et al. 2020; Haug Hilton y Leemann 2020).

El Smartphone, que combina las capacidades de comunicación móvil e Internet, se ha consolidado como una herramienta muy útil para la investigación en ciencias sociales. Los dispositivos, que permiten comunicación tanto sincrónica como asincrónica, se han convertido en artefactos cotidianos que los actores sociales llevan siempre consigo. Además, constituyen la primera forma de acceso a Internet para la mayoría de los usuarios (Kaufmann 2019) y permiten una serie de rutinas diarias, prácticas e interacciones sociales (Ardèvol et al. 2003; Kaufmann 2019; Kaufmann y Peil 2020). Debido a esta omnipresencia de los Smartphones, con sus características distintivas, como la portabilidad, la disponibilidad, las posibilidades de localización y la multimedialidad, ofrecen un amplio potencial metodológico para las ciencias sociales (Schrock 2015). Los Smartphones han desatado un verdadero *boom* como herramientas para recoger grandes cantidades de datos (los así denominados *big data*), por ejemplo, en el campo de la variación lingüística y las *citizen sciences* o *crowd-source sciences* (Haug Hilton y Leemann 2020). Con todo, los teléfonos inteligentes se han empezado a utilizar de manera aún tímida en estudios cualitativos (Kaufmann y Peil 2020; Kaufmann 2019), a pesar de que su naturaleza y diseño interactivo pueden ser altamente potentes para recolectar e incluso generar datos cualitativos (García et al. 2015).

En el siguiente apartado se muestra por qué se decidió incluir entrevistas en un trabajo de investigación cualitativa en el campo de la sociolingüística etnográfica en línea y cómo se desarrolló dicho estudio de caso.

4. Entrevistas mediadas en un estudio de caso

4.1 Descripción del estudio de caso

Este apartado recoge un ejemplo concreto que forma parte de un trabajo de investigación doctoral en sociolingüística etnográfica y en análisis de discurso de tradición enunciativa que incluye entrevistas mediadas como datos de elicitación para su estudio. A continuación, se describe el estudio, desde sus objetivos, el corpus utilizado y la metodología, hasta los cuestionamientos éticos realizados. Todo ello con el fin de comprender el contexto de investigación dentro del cual se llevaron a cabo las entrevistas mediadas, que se presentarán en el apartado 4.2.

4.1.1. Objetivos de investigación y diseño del corpus

Este trabajo cualitativo estudia las ideologías lingüísticas sobre la variación diatópica del español en discursos *epilingüísticos* (en el sentido de Canut 1998, 2000, 2007) de videastas latinoamericanas⁷. Para Cécile Canut, las *producciones epilingüísticas* son aquellas que incluyen a la vez “discursos metalingüísticos en sentido estricto [...] que implican una distanciamiento, un saber y una objetivación con relación al objeto lengua, y los discursos evaluadores espontáneos de los locutores” (1998: 70)⁸. Con el fin de poder

estudiar estas ideologías lingüísticas, se puso en marcha el estudio etnográfico en línea que se describe en el apartado 4.1.2.

El corpus recogido durante este estudio es mixto: consta de veinticinco vídeos extraídos de la plataforma YouTube a propósito de la variación diatópica del español, y de once entrevistas semi-estructuradas con las videastas.

Los vídeos seleccionados duran entre 5 y 15 minutos. Todos ellos fueron producidos entre 2015 y 2021, y pertenecen al género del *vlog* o videoblog, definido entre otros por Christelle Combe (2019: 52) como “un vídeo monologal autocentrado, seguido por comentarios escritos”. Estos vídeos fueron realizados por videastas originarias de Argentina, Chile, Honduras, México, Perú y Venezuela y en ellos las videastas describen las diferencias entre la variedad del español americano y la variedad peninsular de su zona de residencia durante la creación del *vlog*. Estas videastas no son lingüistas, con lo cual este trabajo se interesa por los saberes *profanos* sobre las lenguas y el lenguaje, en el ámbito de la “lingüística popular” o “folklinguistics” (véase Paveau 2008 para un panorama sobre la evolución del campo).

Esta primera parte del corpus se completó con entrevistas semi-estructuradas con las videastas. Desde un punto de vista epistemológico, se trataba de superar el nivel analítico de la materialidad discursiva de los *vlogs* y centrar el foco de interés en las participantes, sus prácticas y lo que estas dicen de ellas. La entrevista, como ya se ha señalado, permite un intercambio directo, extenso y profundo, al contrario de los *vlogs* de YouTube, cuya duración máxima se limita a unos quince minutos y cuyo público es amplio. El objetivo era ahondar en temas clave para el estudio (ideologías, actitudes), retomar elementos citados en los *vlogs* para desarrollarlos y completarlos ya que, como explican Boellstorff et al., “las entrevistas ofrecen la oportunidad de mantener conversaciones verdaderamente privadas que pueden revelar creencias y opiniones a las que es difícil acceder de otro modo” (2012: 93).

4.1.2. Un método de investigación etnográfico

Se eligió un método de investigación etnográfico porque este permite estudiar de manera profunda las prácticas sociales y el sentido que les dan los propios individuos. Además, como afirman Canut et al.,

la inmersión profunda en lugares desconocidos por el investigador o la investigadora [...] sigue siendo una de las mejores maneras de abordar el conjunto de las relaciones sociales, observar los comportamientos, contextualizar las palabras recopiladas, realizar vínculos con los discursos anteriores y exteriores, etc. (2018: 108).

En estos últimos veinte años se han llevado a cabo un gran número de estudios etnográficos centrados en lo digital, que se engloban dentro de un espectro extenso de denominaciones, desde la etnografía virtual (Hine 2000) o para internet (Hine 2015), hasta la etnografía digital (Pink et al. 2016, entre otros), pasando por la netnografía (Kozinets 2010), o la etnografía *en los mundos virtuales* (Boellstorff et al. 2012). Todos estos estudios sociológicos, antropológicos o centrados en los medios de comunicación comparten el enfoque etnográfico, pero difieren en sus principios teóricos y metodológicos. En el marco del presente trabajo, se eligió seguir un método de

investigación etnográfico tradicional, nutriéndolo con reflexiones desarrolladas en los campos citados, con el fin de tomar en cuenta las especificidades del campo digital (al respecto la posición de Boellstorff et al. (2012: 4)⁹).

En efecto, si bien los entornos digitales conllevan especificidades que hace falta tomar en consideración, el método etnográfico es evolutivo y puede adaptarse en función de las disciplinas, los objetivos de investigación y, sobre todo, del tipo de campo, sea *off-line* u *on-line*.

En este sentido, el trabajo de campo se realizó en tres etapas y consistió en una combinación de una inmersión en el campo, seguido por una fase de observación participante en línea, lo que desembocó en las entrevistas mediadas.

De este modo, siguiendo el principio de la etnografía multisituada (Marcus 1995), la inmersión etnográfica empezó durante varios meses en la plataforma de YouTube, y continuó en otras redes sociodigitales afines en las que las videastas se mostraban activas, como Instagram o Facebook. Esta etapa de inmersión permitió conocer el campo, comprender su funcionamiento, sus códigos, y las especificidades de los intercambios que tienen lugar en él. Al final de esta primera fase, se delimitó el corpus principal, que se compone de los veinticinco *vlogs* descritos anteriormente.

Después, se inició la fase de observación participante, durante la cual se creó un canal en YouTube y una cuenta en Instagram, donde se explicaba brevemente en qué consistía el trabajo de investigación con el fin de entrar en contacto con las videastas. La creación del canal se reveló como un método ineficiente porque la plataforma de Youtube no permite acceder con comodidad a los datos a disposición de las videastas ni tampoco ofrece un servicio de mensajería instantánea para entrar en contacto directo con ellas. Dado que todas invitaban a suscribirse a su cuenta de Instagram, se creó un perfil con datos y fotografías personales. Desde este perfil se enviaron peticiones de suscripción a todas las videastas que produjeron los *vlogs*. Además, se comentó en sus publicaciones en Instagram y YouTube, se participó en eventos en línea que ellas organizaban (por ejemplo, unos “en vivo” donde se podía hablar directamente con ellas). En esta fase de observación, se mandaron también mensajes personalizados para iniciar intercambios y solicitar entrevistas.

Seguidamente, se empezó la tercera etapa del proceso, la de las entrevistas mediadas. De las veinticinco videastas contactadas, a veces en varias ocasiones y mediante diferentes medios (mensajería de Instagram, comentarios en YouTube e Instagram, correos electrónicos, Facebook...), más de la mitad contestó (a veces, hasta ocho meses después). Por consiguiente, se pudieron llevar a cabo once entrevistas en total.

4.1.3. Pautas éticas

La metodología de investigación se elaboró tomando en cuenta pautas éticas fundamentales en cualquier estudio que incluya una relación directa o indirecta con personas. En los estudios etnográficos en línea, Hine (2020: 83) precisa que las preocupaciones éticas están muy presentes, aunque cada campo de investigación apela a cuestionamientos específicos (Markham 2006). Desde el principio de este estudio, en la construcción de la metodología, siempre se tomó en cuenta el hecho de que era un trabajo en el cual participaban personas por lo que, como señalan Boellstorff et al., es necesario aplicar “el principio de protección” (*the principle of care*):

El principio de protección surge en parte de las relaciones asimétricas de poder y del desequilibrio de beneficios entre investigador e investigado. El investigador suele ganar mucho más que el informante, obteniendo beneficios que se materializan en trabajo, dinero y reconocimiento profesional. Aunque los informantes pueden disfrutar hablando de sus vidas y su cultura a un investigador, es poco probable que obtengan beneficios a la escala de los etnógrafos. Una consecuencia clave de esta asimetría es el imperativo de que el etnógrafo “cuide bien” a los informantes (Boellstorff et al. 2012: 129-130).

En este sentido, y de manera general en los estudios etnográficos, son fundamentales tanto la protección de las personas como el respeto de su voluntad. En el marco de las entrevistas que se llevaron a cabo en esta investigación, se informó a las participantes de la posibilidad de anonimizarlas antes de empezar. Esta misma información se volvió a dar al final, después de explicarles de qué forma se usarían los datos y qué difusión tendrían. Diez participantes insistieron en que no se anonimizaran ni las entrevistas ni sus vídeos: unas sentían curiosidad por el estudio y querían que apareciera su nombre como huella de su participación, otras explicaban que el estudio las ayudaría a difundir sus producciones incluyendo su nombre y el de su canal¹⁰. Por esas razones, en este artículo aparecen los nombres de algunas de las participantes entrevistadas. En cambio, se anonimizaron los fragmentos de las conversaciones escritas, pues, aunque estaban informadas de que se podían usar sus respuestas en el presente estudio, las capturas de pantalla están sacadas de intercambios privados e informales.

Otro aspecto fundamental que se ha debatido mucho en los estudios etnográficos en línea (Robinson y Schulz 2009; Boellstorff et al. 2012, entre otros) es la manera en la que el investigador o la investigadora debe declarar su presencia en línea. En la Red, es posible llevar a cabo una investigación completa de manera velada, sin que ninguna de las personas observadas se dé cuenta de ello. Este fue el caso de esta investigación durante la primera fase de inmersión exploratoria. Sin embargo, desde la segunda etapa, se puso en marcha una serie de estrategias con el fin de avisar a las participantes de que se estaba investigando sobre ellas y sus prácticas. Se dejó a su disposición todos los datos necesarios para informarlas sobre el estudio con el fin de no investigar *a expensas suyas*, sino *con ellas*. En este sentido, el nombre de la cuenta de Instagram ya indica que se está llevando a cabo un trabajo de investigación. Además, los datos sobre el estudio disponibles en la cuenta lo precisan, así como los primeros mensajes enviados a las participantes a través del servicio de mensajería de Instagram.

Desde esta perspectiva, la entrevista fue útil para asegurarse de que las participantes habían entendido los objetivos del estudio, contestar a sus preguntas, conocer su opinión en cuanto a algunos cuestionamientos éticos y obtener su acuerdo para utilizar y publicar datos y producciones suyas.

4.2. Las entrevistas mediadas desde la génesis hasta los resultados

Este apartado presenta las entrevistas mediadas del estudio de caso. En primer lugar, se exponen las razones que llevaron a realizarlas a distancia; posteriormente, se explica cómo se desarrollaron y, finalmente, se ilustra, mediante dos ejemplos, algunos de los resultados que se observaron.

4.2.1. La decisión de llevar a cabo entrevistas mediadas

En un principio las entrevistas para el estudio estaban previstas para ser desarrolladas cara a cara en España, pero por varias razones de índole contextual se terminaron haciendo a distancia. Primero, la crisis sanitaria de la COVID-19 obligó a las personas a reducir los contactos directos; a partir de la primavera de 2020, en varias ocasiones se cerraron las fronteras entre países, muchos billetes de avión, de tren o incluso de autobús fueron cancelados, y cierta angustia se percibía entre la población debido a la pandemia. Luego, el contexto global en el que se mueven las videastas también llevó a hacer entrevistas a distancia: en el momento en que produjeron los *vlogs*, ellas se encontraban en España, pero algunos años después, cuando se planificaron las entrevistas, muchas ya habían emigrado a otros países o habían regresado a su país de origen. De las once participantes que aceptaron una entrevista, seis se encontraban entonces en España, mientras que las demás estaban en Estados Unidos, México, Chile, Honduras y Países Bajos. En estas circunstancias, no parecía ni ecológico ni económicamente viable desplazarse a todos estos lugares. Finalmente, el contexto digital de una parte del estudio llevó a realizarlas de manera mediada. La inmersión en el campo así como la observación participante ya habían tenido lugar, desde el inicio, en línea. Así pues, las relaciones con las participantes se fueron tejiendo poco a poco de manera asincrónica, esencialmente mediante los servicios de mensajería de Instagram, de correos electrónicos o de comentarios y reacciones a las publicaciones o a los vídeos. En este contexto, parecía más natural utilizar estos mismos servicios para las entrevistas e, incluso, más “seguro” para algunas participantes, como ellas mismas afirmaron en la entrevista (véase apartado 5.2).

4.2.2. Desarrollo de las entrevistas

En su mayoría, las entrevistas semi-estructuradas tuvieron lugar mediante llamada de WhatsApp o videollamada de Instagram, según las preferencias de las participantes (véase apartado 5.2). Su duración media fue de una hora aproximadamente y las temáticas abordadas oscilaban desde las propias trayectorias de las participantes, hasta cuestionamientos relacionados con actitudes e ideologías lingüísticas. Como suele pasar en el caso de entrevistas enmarcadas en estudios etnográficos, las preguntas se apoyaron en las notas de campo tomadas durante la fase de observación, que procedían de las huellas de las videastas en Instagram o YouTube, de los comentarios dejados debajo de los vídeos por otros u otras internautas y de los *vlogs*. Gracias a esta observación previa, se pudieron retomar las categorías utilizadas por las participantes, y hablar con ellas a propósito de fragmentos discursivos apuntados durante esa etapa.

De manera general, las participantes mostraron mucho interés por los temas tratados, haciendo preguntas y animando la conversación; lo cual no es de extrañar puesto que fueron seleccionadas precisamente porque ya habían grabado uno o varios vídeos charlando a propósito de cuestiones lingüísticas. En este sentido, como ya constató Marcón (véase apartado 2), se puede afirmar que la mediación tecnológica no supuso ninguna barrera comunicativa. Las pequeñas diferencias que se percibieron conciernen más bien a las modalidades interlocutivas usadas en la “audiollamada”, donde los marcadores discursivos de retroalimentación requieren ser mucho más numerosos que en las videollamadas o en las entrevistas cara a cara puesto que, debido a la ausencia de contacto visual, estos sirven para constatar y reafirmar la copresencia de las interlocutoras.

4.2.3. Ejemplificación de resultados

A continuación se exponen dos ejemplos que muestran que el empleo de entrevistas mediadas en el marco del presente estudio resultó muy provechoso, ya que permitió, por un lado, ahondar en algunos unos datos epilingüísticos que las videastas solo tocaban superficialmente en sus vídeos y, por otro, aportar información nueva que no se había observado en sus producciones espontáneas.

En primer lugar, para ejemplificar estos resultados, se expondrá el caso de Iliana en relación con el término “caliche catracho”. Iliana es una participante hondureña de unos treinta y cinco años, que estudió periodismo y derecho y vivió durante varios años en Extremadura. En el momento en el que fue contactada se encontraba en Honduras. Durante la preparación de su entrevista, las notas de campo revelaron que, en uno de sus vídeos, ella utilizaba el término “caliche catracho” sin definirlo ni proporcionar ninguna explicación al respecto. La entrevista permitió retomar esta categorización con la videasta y, por tanto, ahondar más en ella:

Fragmento de entrevista n.º 1, participantes: Iliana (ILI) y Nadège (NAD)¹¹

- 1 **NAD** porque tú en tus vídeos/ ((risa)) no he querido buscar lo que era
2 porque prefiero que me lo explicara- que me lo expliques tú
3 **ILI** ajá/
4 **NAD** dices que hablas el (.) caliche catracho [...]
5 **ILI** sí/ ((risa))
6 **NAD** ((risa))
7 **ILI** el caliche es es el caliche es esta parte que por ejemplo para
8 estar en una televisión que no es hondureña pues no se no debería
9 decirlo/ porque son palabras que solo van a entender aquí
10 **NAD** ajá
11 **ILI** es como esa jerga mu:y mu:y autóctona (.) muy de de del país que
12 que no van a entender en ningún lado e:h (.) ningún otro país pues
13 no va a entender más que un hondureño (.) o alguien que ha vivido
14 un tiempo en honduras
15 **NAD** ajá
16 **ILI** entonces a mí en lo personal siempre me gustó (.) usarlo (.) aunque
17 e- en fuera\ me gusta usarlo fuera aunque (.) digamos en: (.) aquí
18 en en honduras (.) tal vez no lo use tanto (.) eh pero fuera me
19 gusta me gustaba más como s- no sé me gustaba hasta como sonaba
20 (.) las palabras estas mu:y mu:y típicas hondureñas
21 **NAD** ajá (.) sí (.) o sea el caliche es esto (.) y catracho significa
22 como del lugar no/
23 **ILI** exacto/ es que a los hondureños el el a los hondureños nos dicen
24 catrachos [...]
25 **NAD** ajá\ y es una palabra: positiva o negativa/ ((risa))
26 **ILI** ((risa)) sí es positiva total sí sí totalmente (.) totalmente
27 positiva
28 **NAD** ((risa))
29 **ILI** nos como nos caracterizamos muchísimo es más (.) e:h ahora salen
30 camisas con catracho catracha eh (.) porque últimamente en los
31 últimos años e- está como muy de moda el (.) el amor patrio (.)
32 el exaltar que eres del país y más cuando estás fuera entonces la
33 gente este compra esas camisas de: de: s- (0.5) por ejemplo de
34 palabras muy muy como digo del caliche mu:y mu:y hondureñas
35 entonces la gente se las compra con las las las camisas con con
36 esas palabras (.) que cualquier hondureño sabe ah/ este es
37 hondureño al habérsela puesto ah este es hondureño\

Este fragmento muestra cómo Iliana proporciona minuciosas explicaciones acerca de la categoría “caliche catracho”, término que tan solo había evocado en uno de sus vídeos. El posicionamiento adoptado por Nadège, doctoranda extranjera del exogrupo, le permite preguntar de manera ingenua por el significado de esta categoría para que Iliana subsane directamente el déficit. Efectivamente, Iliana entiende a través de la indecisión de Nadège, marcada por la micro pausa en la línea 4, que ella no conoce esta categorización, lo que produce la risa en Iliana (l.5) y termina también en risa en Nadège (l.6) para disculparse, como ocurrirá, de nuevo, en las líneas 25 y 28. Es este desconocimiento lo que anima a Iliana a desarrollar el tema y a proporcionar más información sobre el contexto hondureño (l.29-1.<z37), detalles que abren pistas para situar mejor las producciones de la videasta en un contexto general.

En otras ocasiones, las entrevistas han permitido acceder a posicionamientos, actitudes e ideologías que no se hubieran podido conocer de otra forma. En el siguiente fragmento, Rosario, una peruana de unos treinta años, que vive en Ámsterdam en el momento de la entrevista y que ha viajado mucho, menciona, sin que Nadège lo espere, las ventajas de “neutralizar” el español cuando una se encuentra en un contexto de comunicación con hispanohablantes de otras zonas:

Fragmento de entrevista n.º 2, participantes: Rosario (ROS) y Nadège (NAD)

- 1 ROS en cada país la gente habla con sus palabras (.) con con sus cosas
2 NAD ajá
3 ROS y yo desde que empecé a tener amigos internacionales (.) empecé a
4 neutralizar mi español por decirlo así porque yo no hablo como
5 peruana para nada/ para nada\
6 NAD ajá
7 ROS pero esto yo: lo he cogido desde los veinte años (.) desde que:
8 (.) tengo treinta y dos ahorita (.) este: desde que conocí amigos
9 extranjeros porque es que sino no me entienden/
10 NAD ajá
11 ROS es que no puedo habla- hablar como como peruana/ entonces mi mi
12 español yo neutralicé mi español (.) hablo con palabras n- neutras
13 NAD ajá
14 ROS no peruanas
15 NAD sí (.) se entiende claro así es más fácil comunicarse con los
16 demás (.) claro/
17 ROS claro/ solamente hablo como peruana cuando estoy con peruanos (.)
18 pero cuando no hablo neutro\ por ejemplo contigo estoy hablando
19 español neutro
20 NAD ((risa))
21 ROS no te estoy hablando como como como peruana porque posiblemente
22 no me vas a entender\ me dirás eh/ me dirás ((risa))
23 NAD ((risa))

Esa necesidad de “neutralizar” su español, que aborda en varias ocasiones durante la entrevista, aparentemente entra en contradicción con el contenido de sus vídeos, en los que expone las diferencias entre su variedad y la variedad peninsular con la que estaba en contacto en el momento. En este caso, si no se hubiera realizado una entrevista con esta participante, no se habría podido tener acceso a su ideología sobre el español neutro¹² (Sinner 2010 y Arnoux 2020).

5. Recomendaciones prácticas a partir del estudio de caso

Para terminar, después de mostrar cómo se puede trabajar con entrevistas mediadas para estudiar ideologías lingüísticas en el marco de una investigación etnográfica en línea, se proporcionarán una serie de recomendaciones concretas para quienes se dispongan a utilizar este tipo de entrevistas en relación a dos ejes: la creación de una relación de confianza con las personas entrevistadas y la elección del medio y de la modalidad de la entrevista.

5.1. Crear una relación de confianza en un contexto digital

Crear una relación de confianza entre entrevistada y entrevistadora es necesario tanto en los contextos de investigación *online* como *offline* para conseguir entrevistas logradas y productivas. En los entornos digitales, el investigador o la investigadora puede apoyarse en los recursos específicos que tiene a su disposición para ganar la confianza de los y las participantes.

5.1.1. Dejar datos personales y profesionales a disposición de los y las internautas

En el marco de una investigación etnográfica en línea, tanto por razones éticas como para informar a los y las participantes, el investigador o la investigadora puede dejar a disposición de los y las internautas una serie de datos personales y acerca de su trabajo. En este estudio de caso, se apeló a varios recursos: la creación de un canal en YouTube que no dio los resultados esperados por su falta de accesibilidad; la creación de un perfil en Instagram con una presentación escrita breve, limitada a 150 caracteres, y algunas fotos personales de viajes por Latinoamérica; la participación en eventos en línea con las videastas; y la interacción a partir de comentarios dejados bajo sus publicaciones. A pesar de todas estas estrategias, una de las participantes (véase Figura 12) dudó de la veracidad de la cuenta creada, lo que causó ciertas reticencias en ella. Pues, en efecto, como afirma Bernier (2015: 123), “la evaluación” y la “validación social” en las redes sociodigitales depende de una serie de elementos como el número de publicaciones, de seguidores de “likes” en la página, y en este caso la cuenta creada tenía aún muy pocos. Además, Pastinelli (2004) muestra que es frecuente que los y las participantes duden de la veracidad de un trabajo de investigación etnográfico en línea porque no se esperan que los investigadores y las investigadoras se interesen por sus prácticas y producciones. Por eso, se puede considerar también la posibilidad de dejar enlaces a páginas institucionales para evidenciar la afiliación académica, ya que esto puede ser útil en algunos casos.

El siguiente fragmento de entrevista muestra que una participante consultó los datos que se dejaron a su disposición antes de la entrevista, los cuales generaron un sentimiento de confianza en ella. Para contextualizar estos intercambios, Nadège le estaba explicando a Rosario, la participante, que hubiera preferido tener con ella una videollamada en lugar de una llamada simple para así crear más confianza entre ambas. A esta explicación, Rosario le contestó que no era necesario, mediante la repetición del verbo “confiar” en primera persona de singular de presente, puesto que ella ya tenía ese sentimiento de confianza gracias a la presencia de una foto que Nadège puso en su perfil de Instagram:

Fragmento de entrevista n.º 3, participantes: Rosario (ROS) y Nadège (NAD)

1 NAD solamente era (.) a- antes que nada para que vieras (.) quién era
2 yo ((risa)) porque=
3 ROS =((risa)) no pero tranquila confío confío
4 NAD ((risa)) okey=
5 ROS = además (.) t- he visto allí tu foto del salar de uyuni no/
6 NAD sí:::/ estuviste allí/
7 ROS no (.) he ido a: bolivia pero al salar no
8 NAD ajá\ es que claro (.) cuando yo vi tu perfil (.) hace: muy pocos
9 días (.) [e:h]
10 ROS [sí]
11 NAD enseguida pensé que teníamos cosas en común

Así pues, Rosario expone que el hecho de haber visto una foto de un viaje concreto de Nadège a un país cercano al suyo. Tanto este intercambio como la observación formulada en la Figura 12 demuestran, por tanto, que algunas participantes consultan los datos dejados a su disposición y que estos les sirven para evaluar (positiva o negativamente) a la investigadora.

5.1.2. Dedicar atención particular a establecer y mantener los primeros contactos

El esfuerzo dedicado al primer contacto es otro punto fundamental para ganar la confianza de las participantes, pues puede generar tanto una reacción de rechazo como de aceptación en ellas, lo que tendrá consecuencias en el desarrollo posterior de la investigación (Boellstorff et al. 2012: 76). Si bien este hecho no es propio de las entrevistas mediadas, es fundamental para acceder a ellas. En estudios realizados en línea, el contacto se consigue mediante modalidades específicas. En el presente trabajo, los primeros contactos tuvieron lugar en su mayoría de manera asíncrona por Instagram, la Red en la que las participantes se mostraban activas. El primer mensaje tiene que ser claro, breve y transparente, más aún cuando las videastas tienen muchos seguidores, dado que pueden recibir muchas solicitudes a diario. La fase de observación es central, pues permite, por una parte, adaptar el posicionamiento según los códigos y maneras de comunicar en las redes sociales concernidas y, por otra, reunir elementos sobre las personas para personalizar el primer contacto. A continuación, se pueden dejar mensajes de voz mediante algunos servicios de mensajería, como la de Instagram o de WhatsApp, ya que personalizan el intercambio. El envío de un correo electrónico desde la mensajería universitaria puede ser también otra manera de entrar en contacto y, al mismo tiempo, demostrar la afiliación a una institución de prestigio.

5.2. Elegir el medio y modalidad de la entrevista

Este apartado consta de sugerencias en cuanto al tipo de medio elegido para la entrevista, que se fundamentan en las observaciones hechas durante el trabajo de campo, y vienen reforzadas por el análisis de los resultados de una encuesta dirigida a las participantes. Evidentemente, no pretenden ser representativas, sino que apuntan a documentar mejor la percepción que estas tienen de las distintas opciones metodológicas y prácticas que se ofrecen al investigador o a la investigadora, y proporcionan algunas pistas para adaptar eventualmente la metodología en futuros trabajos similares.

5.2.1. La elección de la aplicación o del programa

Para elegir entre los varios programas o aplicaciones que se prestan para una entrevista mediada, es necesario tomar en cuenta las funciones que cada uno de ellos ofrece según los objetivos de investigación y las posibilidades de acceso de los y las participantes. Si bien la aplicación Zoom y su servicio de videoconferencia está ya ampliamente difundida en el mundo académico y presenta ventajas considerables (véase apartado 3) tanto por su facilidad de empleo como por la calidad de su grabación oral o audiovisual, fue ideada para ordenadores y su uso se concentra, sobre todo, en determinados ámbitos laborales y, por lo tanto, menos personales. El teléfono móvil, por su parte, presenta también muchas ventajas en trabajos de investigación similares a este (véase apartado 3), con sus diversas aplicaciones, como WhatsApp, Instagram, o FaceTime, las cuales ofrecen un servicio de llamada y/o videollamada, siempre y cuando los y las participantes posean un Smartphone y una conexión a Internet de calidad.

En el presente estudio, se eligió llevar a cabo las entrevistas mediante las herramientas disponibles en los Smartphones. En un inicio, se quería utilizar por cuestiones de accesibilidad y de homogeneidad una aplicación única, WhatsApp. Sin embargo, algunas participantes se negaron, pidiendo que se usara el servicio de videollamada de Instagram. Por esta razón, se terminó usando ambas aplicaciones para llevar a cabo las entrevistas. Para comprender sus motivos, se propuso a todas las participantes que respondieran una corta encuesta. A este breve cuestionario contestaron nueve de las once participantes.

La primera pregunta se refería a la aplicación elegida para la entrevista: las tres entrevistadas que se negaron a utilizar WhatsApp lo justificaron por razones de seguridad (Figura 1 y Figura 2) o de accesibilidad (Figura 3), como se puede constatar en las siguientes capturas de pantalla de sus respuestas por WhatsApp o Instagram desde un Smartphone:



Figura 1

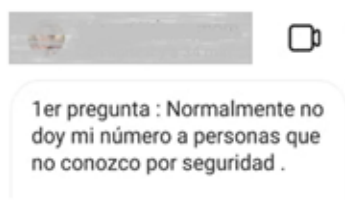


Figura 2



Figura 3

Se entiende con estas respuestas que conversar por WhatsApp implica un grado de cercanía mayor, materializado por el intercambio de los números de teléfono personales. Estas participantes, por tanto, se sentían más protegidas hablando mediante la cuenta de una red sociodigital, que ya saben pública, que teniendo que proporcionar un dato personal más, su número de teléfono.

5.2.2. La modalidad de la entrevista

La elección de la modalidad del intercambio (audio, audiovisual o presencial) también varía en función de los objetivos del estudio, del tipo de análisis que se quiere llevar a

cabos, del contexto, de las condiciones del estudio y de la voluntad de los y las participantes.

En este trabajo, se partió de la hipótesis de que la entrevista cara a cara hubiera sido la mejor opción, seguida por las videollamadas, ya que ambas modalidades suponían en teoría un grado de proximidad mayor entre las interlocutoras. No obstante, algunas participantes que usaban Whatsapp prefirieron una llamada simple a una videollamada (la cuestión no se planteaba con Instagram, que no ofrece la posibilidad de una llamada simple). La encuesta permitió comprobar las preferencias de cada una, y las razones de su elección, que fueron variadas, como lo muestran las Figuras de la 4 a la 7:



Figura 4

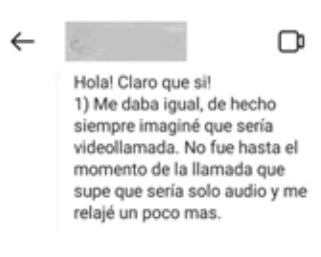


Figura 5

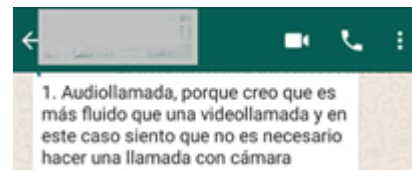


Figura 6



Figura 7

Aunque se aprecien las dos modalidades, se constata una preferencia por la audiollamada: una participante dio a entender que la videollamada provoca más presión (Figura 5), otra aludió a su fluidez (Figura 6), otra mencionó en un mensaje su practicidad, explicando que requiere menos organización que la videollamada ya que se puede hacer incluso desde la calle o en el trabajo.

La última pregunta del cuestionario apuntaba a comprobar si, en el caso en el que hubiera sido posible, hubieran aceptado una entrevista cara a cara. Siete participantes contestaron positivamente, tres con la condición de que se hubiera desarrollado en un lugar público, de las cuales, dos mencionan razones de seguridad para justificar esta condición (Figura 8 y Figura 9). La falta de confianza también aparece en las respuestas de la Figura 12 y Figura 13, cuyas participantes seguramente hubieran rechazado una entrevista cara a cara en un primer momento, aunque sí la hubieran aceptado tras la entrevista mediada. La entrevista mediada se ve, pues, como una etapa en la relación, puesto que ha permitido alcanzar un grado de confianza mayor:

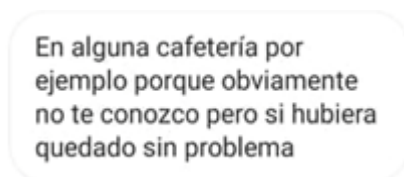


Figura 8

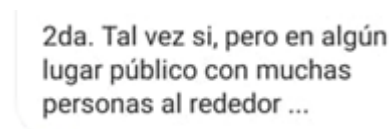


Figura 9

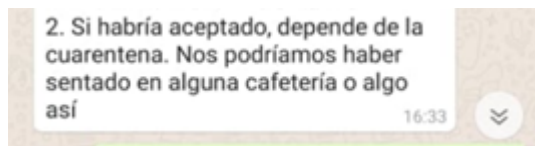


Figura 10

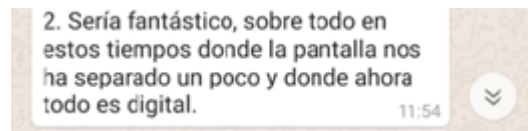


Figura 11

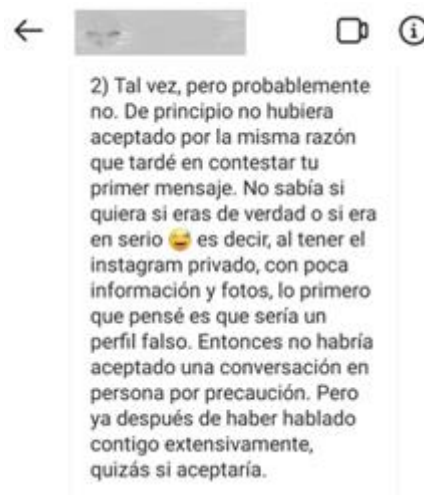


Figura 12

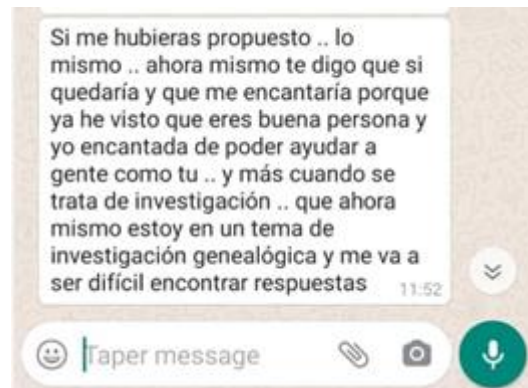


Figura 13

En este punto se percibe que el carácter mediado de la entrevista puede ser una ventaja en este tipo de estudios, ya que puede promover un sentimiento de seguridad en las participantes en comparación con la entrevista cara a cara.

Para terminar, se observa que el contexto sanitario actual también pudo tener una incidencia en las respuestas de las participantes: se mencionó en dos ocasiones, una como posible impedimento para una entrevista cara a cara (Figura 10) y otra, al contrario, como factor que refuerza la voluntad de verse en persona (Figura 11).

En suma, las observaciones y recomendaciones formuladas en esta parte no pretenden tener un alcance generalizante. Apuntan más bien a destacar algunos procedimientos y cuestionamientos que resultaron provechosos en el marco del presente trabajo con el fin de documentarlo con vistas a futuros estudios similares. La temática central que guía este apartado concierne a las estrategias desarrolladas para alimentar el sentimiento de confianza mutua entre participante e investigador o investigadora para obtener entrevistas mediadas logradas en el marco de un estudio etnográfico en línea. En este sentido, se recomienda dedicar un esfuerzo particular a la presentación en línea de la investigadora o del investigador, y de su trabajo de investigación, ya sea mediante las distintas herramientas a disposición, o en los primeros contactos e intercambios asíncronos, que, como se ha expuesto, son determinantes. Luego, se ha mostrado la importancia de realizar encuestas posteriores a la entrevista mediada para poder documentar la autopercepción de las entrevistadas con respecto tanto al desarrollo de la entrevista como a la elección del medio y los dispositivos empleados. Esto ha permitido tomar en cuenta la opinión de las participantes, con el fin de mejorar sus sentimientos de seguridad y de confianza, y

poder, en el futuro, ajustar la metodología en trabajos similares. Pero más allá de estos aspectos específicos, lo que se ha querido mostrar con estos ejemplos es que para poder acceder a datos en un terreno etnográfico en línea es muy importante ser flexibles y adaptarse a los requerimientos de las y los participantes, siempre y cuando estos no mellen los objetivos fundamentales del estudio (Levon 2013: 71).

6. Conclusión

En este artículo se ha mostrado que la entrevista mediada puede ser una herramienta eficiente para quien investiga sobre ideologías y actitudes lingüísticas, con tal de que esté bien integrada en el diseño metodológico de todo el estudio y vaya en consonancia con él. Las entrevistas mediadas incluidas en este estudio fueron extensas, productivas y permitieron cumplir con los objetivos iniciales: ahondar en temas que las participantes solo habían mencionado en sus *vlogs* y aportar elementos nuevos en relación con las temáticas que interesaban.

Sin embargo, para que la entrevista mediada pueda cumplir su función, hace falta tomar en cuenta una serie de cuestiones. Primero, la distancia física que supone la entrevista mediada no conlleva una descorporeización de la misma. En este sentido, la presentación de la investigadora o del investigador, así como sus primeros contactos con las y los participantes son fundamentales puesto que de ellos depende la posibilidad de llevar a cabo una entrevista. Por esta razón, estos aspectos deben plantearse y trabajarse desde un inicio en el diseño metodológico (véase Ardèbol 2003). Además, en la organización de la entrevista, el investigador o la investigadora tiene que adaptarse tomando en cuenta tanto las precauciones y los temores de los o las participantes, como su voluntad en la elección del medio o de la aplicación utilizada. Precisamente, por eso, dispositivos como los Smartphones ofrecen mayor flexibilidad que los ordenadores para estudios como el nuestro, puesto que, como indican Kaufmann y Peil (2020: 230), permiten la comunicación inmediata entre investigadores e investigadoras y participantes, proporcionan una herramienta fácil de usar para establecer el contacto y dar pie a las entrevistas e incluyen varias opciones para la comunicación, como el texto escrito, las imágenes y la grabación en vídeo.

A nivel ético, no se puede olvidar que los datos en línea no están a libre disposición del investigador o de la investigadora para los usos que él o ella quiera: declarar su presencia y trabajar con el acuerdo de las personas involucradas toda vez que sea posible es fundamental. Cada paso metodológico se tiene que planificar pensando en la necesidad de proteger a los y las participantes y de respetar su voluntad.

Para terminar, es importante documentar y estudiar más las experiencias de los y las participantes en los procesos de entrevistas mediadas para mejorar su calidad y explorar nuevas formas de aplicación. Por eso, resulta de provecho documentar las impresiones y subjetividades de estos actores y estas actoras sociales mediante encuestas que pueden enviarse después de la entrevista (Archibald et al. 2019; Leeman et al. 2020; Kaufmann y Peil 2020). De esta forma, se puede mejorar las prácticas de investigación de manera colaborativa, incluyendo a las personas con las que se ha trabajado.

7. Referencias bibliográficas

- Archibald, Mandy M.; Ambagtsheer, Rachel C.; Casey, Mavourneen G.; Lawless, Michael. 2019. Using Zoom videoconferencing for qualitative data collection: Perceptions and experiences of researchers and participants. *International Journal of Qualitative Methods* 18: 1-8.
- Ardèvol, Elisenda; Bertrán, Marta; Callén, Blanca; Pérez, Carmen. 2003. Etnografía virtualizada: la observación participante y la entrevista semiestructurada en línea. *Athenea Digital. Revista de pensamiento e investigación social* 3: 72-92.
- Arnoux, Elvira Narvaja de. 2020. De la “unidad en la diversidad” al “español auxiliar internacional”. En S. Greußlich y F. Lebsanft, eds. *El español, lengua pluricéntrica. Discurso, gramática, léxico y medios de comunicación masiva*. Göttingen: V&R unipress / Bonn University Press, pp. 39-60.
- Bailey, Guy; Bernstein, Cynthia. 1989. Methodology of a phonological survey of Texas. *Journal of English Linguistics* 22: 6-16.
- Bailey, Guy; Dyer, Margie. 1992. An approach to sampling in dialectology. *American Speech* 67: 3-20.
- Bailey, Guy; Wikle, Thomasan; Tillery, Jan. 1997. The effects of methods on results in dialectology. *English World-Wide* 18: 35-63. DOI: 10.1075/eww.18.1.03bai
- Becker, Kara. 2013. The sociolinguistic interview. En C. Mallinson, B. Childs y G. Van Herk, eds. *Data collection in sociolinguistics: Methods and applications*. Londres: Routledge, pp. 99-100.
- Bernier, Guylain. 2015. *La vidéo de soi sur internet : rendre visible sa différence. Au-delà de la technologie, les fondements sociaux*. Paris: L’Harmattan.
- Boellstorff, Tom; Nardi, Bonnie; Pearce, Celia; Taylor, T.L. 2012. *Ethnography and virtual worlds. A handbook of method*. Princeton: Princeton University Press.
- Briggs, Charles L. 1986. *Learning how to ask: a sociolinguistic appraisal of the role of the interview in social science research*. Cambridge: Cambridge University Press.
- Bürki, Yvette. 2020. Connecting micro and macro sociolinguistic processes through narratives. A glottopolitical Gaze. *Journal of multilingual and multicultural development* 41.1: 12-24.
- Canut, Cécile; Danos, Félix; Him-Aquilli, Manon; Panis, Caroline. 2018. *Le langage, une pratique sociale. Eléments d’une sociolinguistique politique*. Besanzón: Presses universitaires de Franche-Comté.
- Canut, Cécile. 1998. Pour une analyse des productions épilinguistiques. *Cahiers de praxématique* 31: 69-90.
- Canut, Cécile. 2000. Subjectivité, imaginaires et fantasmes des langues: la mise en discours “épilinguistique”. *Langage et société* 93: 71-97.
- Canut, Cécile. 2007. L’épilinguistique en question. En G. Siouffi y A. Steuckardt, eds. *Les linguistes et la norme*. Berlin: Peter Lang, pp. 49-72
- Cater, Janet. 2011. SKYPE—A cost-effective method for qualitative research. *Rehabilitation Counselors & Educators Journal* 4: 10-17.
- Codó, Eva. 2008. Interviews and questionnaires. En L. Wei y M. G. Moyer, eds. *The Blackwell guide to research methods in bilingualism and multilingualism*. Londres: Blackwell, pp.158-176.
- Combe, Christelle. 2019. Les genres numériques de la relation. *Langage et société* 167: 51-80.

- Davis, Boyd. 2013. Other interviewing techniques in sociolinguistics. En C. Mallinson, B. Childs y G. Van Herk, eds. *Data collection in sociolinguistics: Methods and applications*. Londres: Routledge, pp. 114-117.
- De Fina, Anna; Perrino, Sabina. 2011. Introduction: Interviews vs. natural contexts: A false dilemma. *Language in society*, 40.1: 1-11.
- Ellingston, Laura E. 2012. Interview as embodied communication. En J. F. Gubrium, J. A. Holstein, A. B. Marvasti y K. D. MacKinney, eds. *The SAGE handbook of interview research: The complexity of the craft*. Thousand Oaks: Sage, pp. 525-539.
- Feagin, Crawford. 2013. Entering the Community. En J. K. Chambers, P. Trudgill y N. Schilling-Estes, eds. *The handbook of language variation and change*. Oxford: Wiley-Blackwell, pp. 17-37.
- Fontana, Andrea; Frei, James H. 1994. Interviewing: the art of science. En N. K. Denzine y Y. S. Lincoln, eds. *The handbook of qualitative research*. Thousand Oaks: Sage, pp. 361-376.
- García, Borja; Welford, Joanna; Smith, Brett. 2015. Using a smartphone app in qualitative research: The good, the bad and the ugly. *Qualitative Research* 16.5: 508-525.
- García, Ofelia; Flores, Nelson; Spotty Massimiliano, eds. 2017. *The Oxford handbook of language and society*. New York: Oxford University Press.
- Gibson, Barbara; Hua, Zhu. 2016. Interviews. En Z. Hua, ed. *Research Methods in Intercultural Communication. A practical guide*. New York: John Wiley and Sons, pp. 182-195.
- Gómez Font, Alberto. 2012. Español neutro o internacional.
www.fundeu.es/escribireninternet/espanol-neutro-o-internacional/
- Haug Hilton, Nanna; Leemann, Adrian. 2020. Editorial: using smartphones to collect linguistic data. *Linguistic Vanguard* 2021 7(s1): 1-7.
- Hernández Campoy, Juan Manuel; Almeida, Manuel. 2005. *Metodología de la investigación sociolingüística*. Granada: Comares.
- Herold, Ruth. 1990. *The implementation and distribution of the low back merger in eastern Pennsylvania*. Ph.D. dissertation, University of Pennsylvania.
- Hine, Christine. 2000. *Virtual ethnography*. Londres: Sage.
- Hine, Christine. 2015. *Ethnography for the internet: Embedded, embodied and everyday*. Londres: Bloomsbury.
- Hine, Christine. 2020. L'ethnographie des communautés en ligne et des médias sociaux: modalités, diversités, potentialités. En M. Millette, F. Millerand, D. Myles, G. Latzko-Toth, eds. *Méthodes de recherche en contexte numérique*. Montreal: Presses de l'Université de Montréal, pp. 77-101.
- James, Nalita; Busher, Hugh. 2012. Internet interviewing. En J. F. Gubrium, J. A. Holstein, A. B. Marvasti y K. D. MacKinney, eds. *The SAGE handbook of interview research: The complexity of the craft*. Thousand Oaks: Sage, pp. 177-192.
- Kaufmann, Katja. 2019. Mobile methods: Doing migration research with the help of smartphones. En K. Smets, K. Leurs, M. Georgiou, S. Witteborn, y R. Gajjala, eds. *The SAGE Handbook of Media and Migration*. Sage Publications: Londres, pp. 167-179.
- Kaufmann, Katja; Peil, Corinna. 2020. The mobile instant messaging interview (MIMI): Using WhatsApp to enhance self-reporting and explore media usage in situ. *Mobile Media & Communication* 8.2: 229-246.
- Kozinets, Robert V. 2010. *Netnography*. Londres: Sage.

- Labov, William. 1966. *The social stratification of English in New York City*. Washington: Center for Applied Linguistics.
- Leemann, Adrian; Jeszenszky, Péter; Steiner, Carina; Studerus, Melanie; Messerli, Jan. 2020. Linguistic fieldwork in a pandemic: Supervised data collection combining smartphone recordings and videoconferencing. *Linguistics Vanguard*, 6.s3: 20200061.
- Levon, Erez. 2013. Ethnographic Fieldwork. En C. Mallinson, B. Childs y G. Van Herk, eds. *Data Collection in Sociolinguistics. Methods and Applications*. New York: Routledge, pp. 69-79.
- Lo Iacono, Valeria; Symonds, Paul; Brown, David. H. 2016. Skype as a tool for qualitative research interviews. *Sociological research online*, 21.2: 103-117.
- Marcón, Osvaldo A. 2021. Las entrevistas a distancia en Trabajo Social Forense: reflexiones teórico-prácticas. *Itinerarios de Trabajo Social* 1: 87-94.
- Marcus, George E. 1995. Ethnography in/of the world system: The emergence of multi-sited ethnography. *Annual review of anthropology* 24.1: 95-117.
- Markham, Annette N. 2006. Ethic as method, method as ethic: A case for reflexivity in qualitative ICT research. *Journal of Information Ethics* 15.2: 7-54.
- Márquez Reiter, Rosina. 2018. Interviews as sites of ideological work. *Spanish in Context* 15.1: 54-76.
- Moreno Fernández, Francisco. 2009. *Principios de sociolingüística y sociología del lenguaje*. Barcelona: Grupo Planeta (GBS).
- Pastinelli, Madeleine. 2004. Les limites floues de l'ethnologie du contemporain: quelques réflexions autour d'une enquête sur la sociabilité électronique. *Ethnologies* 26.2: 221-255.
- Paveau, Marie-Anne. 2008. Les non-linguistes font-ils de la linguistique? *Pratiques* 139-140: 93-109. <https://journals.openedition.org/pratiques/1200>
- Pink, Sarah; Horst, Heather; Postill, John; Hjorth, Larissa; Lewis, Tania; Tacchi, Jo, 2016. *Digital ethnography. Principles and practice*. Londres: Sage.
- Rigatuso, Elizabeth M. 2019. En torno a fenómenos de variación pragmática regional en interacciones de servicio comerciales del español bonaerense: los marcadores interactivos: a propósito de dale. *Revista Internacional de Lingüística Iberoamericana* 33.1: 105-130.
- Robinson, Laura; Schulz, Jeremy. 2009. New avenues for sociological inquiry: Evolving forms of ethnographic practice. *Sociology* 43.4: 685-698.
- Schrock, Andrew Richard. 2015. Communicative affordances of mobile media: Portability, availability, locatability, and multimediality. *International Journal of Communication* 9: 1229-1246.
- Sinner, Carsten. 2010. ¿Es neutro el español neutro? En M. Iliescu, H. Siller-Runggaldier, P. Danler (dirs.), *Actes du XXVe Congrès International de Linguistique et Philologie Romanes*, vol. 3, Berlin, de Gruyter, p. 707-716.
- Vida Castro, Matilde Ángeles; Ávila Muñoz, Antonio Manuel; Carriscondo Esquivel, Francisco Manuel. 2016. *Manual práctico de sociolingüística*. Madrid: Síntesis
- Wardhaugh, Ronald; Fuller, Janet. 2021. An introduction to sociolinguistic. Hogoboken, New Jersey: Wiley Blackwell.
- Weller, Kathrin. 2015. Accepting the challenges of social media research. *Online Information Review* 39.3: 281-289.

¹ Sobre el desarrollo de la sociolingüística desde sus inicios variacionistas hasta la actualidad, que incluye varias vertientes epistemológica y metodológicamente diversas, véanse Wardhaud y Fuller 2021, y García et al. 2017.

² Lamentablemente, aún no hay suficientes trabajos sobre la aplicación de entrevistas mediadas por Internet en sociolingüística.

³ Puede consultarse la información sobre el proyecto en

https://www.ling.upenn.edu/phono_atlas/home.html (última fecha de consulta 10.03.2021).

⁴ Tampoco los traen los manuales consultados (Hernández Campoy y Almeida 2005, Vida Castro et al. 2016 y Moreno Fernández 2009).

⁵ Cifras según las estadísticas de Internet World Stats

(<https://www.internetworldstats.com/emarketing.htm>, fecha de consulta: 15.03.2021).

⁶ Sobre los primeros empleos en las ciencias sociales de entrevistas electrónicas en Internet, de naturaleza asincrónica, véase James y Busher (2012).

⁷ Se utiliza el femenino inclusivo para designar a las personas de género masculino y femenino que participaron en el estudio, dado que la mayoría son mujeres.

⁸ Todas las traducciones de citas del francés, o del inglés al español son nuestras.

⁹ Al igual que muchos académicos, en ocasiones utilizamos conversando términos como “etnografía digital”, “etnografía virtual” o “etnografía de Internet”. Sin embargo, consideramos que estas etiquetas son equívocas en general, ya que la metodología etnográfica se traslada con elegancia y fluidez a los mundos virtuales. Nos vemos como etnógrafos que investigan en mundos virtuales, no como “etnógrafos virtuales”. (Boellstorff et al. 2012: 4)

¹⁰ Se eligió respetar su voluntad, lo cual implica preguntarse cada vez que se cita un fragmento de estas entrevistas si podrían poner en peligro a las participantes, en cuyo caso se anonimizarían.

¹¹ Convenciones de transcripción ICOR del laboratorio ICAR Lyon 2 y ENS Lyon (UMR 5191):

(.) pausa corta

[] superposición

/ entonación ascendente

\ entonación descendente

: sonido alargado

-palabra no terminada

(()) indicaciones paraverbales o no verbales

¹² El concepto de *español neutro* surge en los años 70 del siglo pasado como una variedad normalizada artificialmente y desterritorializada que sirva como solución para acceder comercialmente a un público hispánico de procedencias nacionales diversas y que maneja diferentes variedades diatópicas del español en la expansión de la industria audiovisual y la puesta en marcha de medios de comunicación internacionales (Arnoux 2020: 47). Gómez Font, uno de los coordinadores más influyentes del proyecto Fundeu (Fundación del Español Urgente), denominaba esta variedad comercial del español como “español común, neutro, internacional” (Gómez Font 2012: 22-23).

Claudia Crespo del Río (Pontificia Universidad Católica del Perú). Reseña de Fuller, Janet M.; Leeman, Jennifer. 2020. *Speaking Spanish in the US. The Sociopolitics of Language*. Bristol: Multilingual Matters.

Este libro es una publicación sumamente interesante para aprender y reflexionar acerca del contexto social, político e ideológico del español en los Estados Unidos en la actualidad y también acerca de las condiciones históricas y otros antecedentes que han moldeado esa situación. Además, no solo se enfoca en la lengua española como elemento central, sino principalmente en la relación con sus hablantes e incluso en la común vinculación que se establece entre esta lengua y las comunidades latinas en el país. Para realizar una apreciación más precisa de la publicación, esta reseña está organizada como sigue: la primera sección se ocupa del capítulo introductorio del libro por su relevancia para la obra en general, la segunda sección analiza su propósito principal y la relación con el marco teórico amplio que desarrolla, la tercera sección apunta a comentar sobre la organización de la obra y su utilidad para el público lector, la cuarta sección presenta de manera ordenada el contenido de cada capítulo y añade algunas evaluaciones específicas. Finalmente, la reseña se cierra con una apreciación final desde una perspectiva general.

1. Sobre la introducción

La obra es presentada como una segunda edición a la que se le han realizado cambios notorios, como el título y la inclusión de nuevos capítulos. No es el objetivo de esta reseña presentar una comparación entre las dos ediciones, pero sí es importante destacar el acierto de las autoras de hacer explícitos, con mucho detalle, los cambios de esta nueva edición. De esta manera, el capítulo 1, “An Introduction to Speaking Spanish in the US”, desarrolla una sección que los presenta y justifica; como ejemplo, vale la pena subrayar el que el nuevo título incluya la frase *The sociopolitics of language*, lo que enfatiza el enfoque profundamente social que se busca darle al libro. El interés principal de Fuller y Leeman es abordar el estudio del lenguaje como acción social, de modo que el español en los Estados Unidos constituye un caso idóneo para desarrollar y aplicar los conceptos teóricos involucrados en un estudio sociopolítico del lenguaje (p. 2). En efecto, sobre la base de esta propuesta, los cambios introducidos en esta segunda edición responden al objetivo de analizar el caso del español desde una perspectiva social y política.

Además de explicar las novedades de la edición, el capítulo 1 cumple la función de marco general para establecer las bases teóricas sobre las que se apoya el libro. Por esa razón, se deja en claro que se desarrolla un enfoque interdisciplinario para analizar la relación entre el lenguaje y el mundo social y político: “Just as language use is inherently social and political, the social and political world is also shaped by language” (p. 3). Los principales conceptos teóricos que se presentan para sustentar ese enfoque son el construccionismo social y los estudios críticos, como el Análisis Crítico del Discurso, que orientan el análisis hacia el estudio de las ideologías. De hecho, las ideologías lingüísticas son una noción fundamental a lo largo del libro, no solo para entender cuáles de estas predominan en los diversos contextos de uso del español en los Estados Unidos, sino también para proponer las motivaciones detrás de las valoraciones, creencias y prácticas sociales vinculadas al español en ese país. Finalmente, el capítulo

incluye una justificación sobre por qué se usan ciertos términos para los grupos e individuos de los que más se habla en el libro, como *latinxs*¹ y *anglos*. Las autoras sostienen que, en el caso del primer grupo, el término puede variar a lo largo de los capítulos dependiendo de si se hace referencia a estudios en los que se utilicen otras opciones.

2. Sobre el propósito principal

El capítulo 1 se cierra con la descripción breve de cada capítulo (y la alusión a si es uno nuevo con respecto a la primera edición). Antes de comentar lo que cada uno de estos capítulos presenta, es indispensable reparar en la intención de las autoras de hacer explícita su postura con respecto al estudio del español en los Estados Unidos: asumen un posicionamiento crítico frente a los asuntos lingüísticos, históricos y sociales que giran en torno a hablar español. Por ello, prima el análisis de ideologías lingüísticas y de representaciones sociales como eje transversal de toda la obra. Así, Fuller y Leeman se encargan de hacerle recordar al público lector nociones teóricas vistas en capítulos previos cuando presentan algún caso en particular. Esto es muy común, por ejemplo, cuando traen a colación una ideología lingüística específica que ya ha sido explicada con anterioridad; es el caso de la ideología de una nación-una lengua (*one nation-one language ideology*), que se explica en el capítulo 4, correspondiente a ideologías lingüísticas (pp. 76-80), y luego es evocada en el capítulo sobre políticas lingüísticas (capítulo 8) como sustento teórico para analizar las políticas lingüísticas que buscan que exista una lengua específica como idioma oficial de un país; esta ideología se invoca para entender las motivaciones de aquellos grupos y organizaciones que promueven el inglés como única lengua oficial en los Estados Unidos (p. 193). Otro momento en que se recuerda esta ideología es en el capítulo acerca de la enseñanza y el uso del español en las escuelas (capítulo 9), puesto que se plantea que es una de las razones por las que los programas de enseñanza que apuntan al mantenimiento y el bilingüismo no son los preferidos para ser financiados a nivel federal o estatal (p. 210). Este recuerdo constante permite procesar mejor la relación entre diferentes dimensiones que no necesariamente parecen estar vinculadas. Mediante esta estrategia, el libro consigue una estructura mucho más concisa, en la que es posible inferir que las situaciones en las que está involucrada la lengua española, sus hablantes o las comunidades latinas no son únicamente reflejo de características lingüísticas, sino que pueden ser analizadas también mediante herramientas sociales, históricas e ideológicas.

Este es el punto más fuerte de la publicación, ya que se hace sumamente necesario contar con libros que elaboren una argumentación que demuestre cómo las representaciones de una lengua y de sus hablantes no son objetivas, arbitrarias o producto del azar. En las sociedades actuales, es frecuente encontrar reacciones adversas ante estas propuestas, lo que hace evidente que el orden impuesto por grupos hegemónicos se asume como lo normal o lo indiscutible, tal como sostiene Burr (2013). Esta autora aborda el concepto de construccionismo social para argumentar que aquello que generalmente no se cuestiona, que se da por sentado, es en realidad producto de la estructura social, y es esta también la que se encarga de mantener y reforzar esas ideas. Así, por ejemplo, la personalidad, según la perspectiva construccionista, debe entenderse como moldeada por aspectos culturales, históricos y sociales, y no como un rasgo esencial de la naturaleza humana. De la misma forma, los juicios morales acerca de las prácticas sexuales de las personas están cercanamente conectados con la organización de la sociedad en que viven; prueba de ello es que otro tipo de prácticas, como los gustos

por cierto tipo de comida o bebida, no producen las mismas reacciones de exaltación. Sin embargo, los miembros de una sociedad han interiorizado tan profundamente esas creencias que las asumen como naturales, con lo que el sustento teórico del construccionismo social resulta contraintuitivo y resistido por muchas personas.

En cuanto a las construcciones sociales acerca del lenguaje, podemos encontrar en el libro de Fuller y Leeman ejemplos concretos en que las personas o los grupos las naturalizan. En el contexto estadounidense, se cita el caso de la hegemonía del inglés en las escuelas: otras lenguas son vistas como obstáculos para el rendimiento académico de los estudiantes, de manera que se prioriza el aprendizaje del inglés para quienes las hablan y no se presta atención a los avances que tengan en otras materias (p. 213). Al contrario de conceder la razón a la idea de que hablar otras lenguas afecta el rendimiento escolar, las autoras refieren a especialistas que lo explican como un caso de racialización, en que al español se le adjudica un valor indexical de latinidad. Así, no se evidencia directamente la jerarquización social en la que los grupos latinos se ubican en una posición de opresión, sino que esta se camufla por medio de otros factores que son normalizados por la sociedad, como el asumir que el inglés debe ser la única lengua de instrucción escolar. Muchas veces estas ideologías lingüísticas no son vistas de manera clara por los miembros de la sociedad, en especial por quienes forman parte de grupos hegemónicos. Fuller y Leeman buscan así demostrar detalladamente por qué sí hay una asociación entre lenguaje e ideas racializadoras.

3. Sobre la organización

Contar con publicaciones que tengan como uno de sus principales objetivos demostrar un vínculo tan contundente entre las creencias sobre las lenguas y las representaciones de otros aspectos sociales es valioso y mucho más si están organizadas en un formato que se orienta hacia estudiantes. Eso es destacable en este libro, puesto que su estructura lo hace ideal para ser usado como un libro de texto en clases universitarias, incluso de primeros niveles, o como un autoinstructivo para quien tiene el interés de aprender sobre la situación del español desde su propia iniciativa.

De este modo, si evaluamos la estructura de cada capítulo, observamos que cada uno mantiene un formato uniforme que resulta muy didáctico: al inicio, se presentan explícitamente los objetivos que se buscan lograr en el capítulo; luego, se desarrolla una breve sección de introducción, en la que se crea una conexión evidente entre el tema general del capítulo y otros temas ya desarrollados hasta ese momento, además de adelantar cuáles serán las ideas principales que se trabajarán. De manera similar, se termina con una sección de conclusiones llamada “Conclusions and connections”, debido a que no solo se elabora una breve síntesis de los rasgos principales, sino que también se los asocia con lo ya visto en capítulos anteriores e incluso con aquellos que se desarrollarán posteriormente. Finalmente, se incluye una sección de preguntas de discusión y actividades, y una lista de referencias bibliográficas adicionales a las mencionadas a lo largo del capítulo. La sección de preguntas de discusión y actividades es sumamente provechosa para que el público lector pueda ampliar sus ejemplos y también para que pueda reflexionar sobre su propia experiencia al respecto. Esto resulta interesante, porque el público principal al que está dirigida la obra son hablantes de inglés y de español que conocen de cerca el contexto estadounidense –aunque la versión en español de la obra se publicará pronto y con ello se expandirá su recepción–, por lo que el libro les invita a cuestionar sus propias creencias y las de su entorno más cercano

al proponerles dialogar e indagar acerca de estas. Para ello, las actividades se valen de recursos diversos, tales como los estudios detallados en el capítulo, videos de medios de comunicación y redes sociales, artículos periodísticos, *podcasts*, etc.

Con respecto a la bibliografía adicional, con excepción del último capítulo, se ofrece una lista de otros estudios relevantes además de los que ya se han mencionado a lo largo del desarrollo. No obstante, no debe pasar desapercibido el que la mayoría de estos estudios no hayan sido incluidos en la lista general de referencias bibliográficas que se encuentra al final de libro. Habría sido adecuado incluirlos también, de modo que se contara con una sección que englobara todas las referencias, o si no, crear otra lista de referencias al final en la que se precisara que se trataba de esas lecturas adicionales sugeridas en cada capítulo. Esto podría evitar la confusión, ya que hay varios casos en que algunos de estos textos adicionales sí son mencionados en las secciones de desarrollo del capítulo y, por lo tanto, incluidos en la lista general de referencias. Dos ejemplos son las publicaciones de Dávila (2012) en el capítulo sobre la presencia del español en medios de comunicación (capítulo 7) y Valenzuela (1999) en el capítulo sobre la enseñanza y el uso del español en las escuelas (capítulo 9).

Como ya se mencionó, la parte final del libro contiene una lista general de referencias, pero también un glosario de los términos más importantes y un índice temático, bastante útil considerando que las autoras recurren a la constante conexión de conceptos entre capítulos. Es rescatable que los términos definidos en el glosario final aparezcan resaltados dentro de los capítulos, de forma que se recuerde al público lector la posibilidad de consultar la definición propuesta al final del texto. Un comentario al margen sobre esta opción es que no queda suficientemente claro cómo se definió cuál de las veces en que se usa el término sería la resaltada, ya que no parece ser necesariamente la primera o aquella en que aparenta ser más relevante.

4. Sobre el contenido de los capítulos

Antes de comentar el contenido de cada capítulo, es necesario detenernos en la idea de que el libro está dirigido a un público de estudiantes. De esta forma, se debe tomar en cuenta que los ejemplos y casos presentados no buscan necesariamente parecerse a las experiencias propias de quien lee, puesto que es probable que la familiaridad de algunas personas con el español o con grupos latinos sea menor. Por esa razón, las explicaciones a lo largo de los capítulos recurren de manera constante al análisis de situaciones vividas por otros grupos minoritarios en los Estados Unidos. Así, es frecuente encontrar alusiones al uso y las creencias acerca de variedades de inglés o de lenguas habladas por grupos afroamericanos y grupos nativos americanos. No resulta muy distante considerar casos como estos para explicar lo que ocurre con el español en los Estados Unidos, debido a que también se trata de grupos racializados. En ese sentido, es posible identificar temáticas cercanas a la lengua afroamericana o *African American Language* (AAL)²: King (2020), por ejemplo, discute los principales criterios que se hacen necesarios en los estudios lingüísticos sobre esta lengua, como la manera en que se la denomina, la construcción de identidades, la esencialización en la que se puede caer al establecer correspondencias entre una variedad y un grupo social, entre otros. Una revisión similar se puede encontrar en el libro de Fuller y Leeman, como en el capítulo 10, que se centra en las características estructurales del español en los Estados Unidos: las autoras comentan los conceptos de *codeswitching* o cambio de código, *translanguaging* o translenguando, e incluso el de lengua, y presentan los sustentos que

especialistas han elaborado con respecto a preferir unos términos sobre otros. También, elaboran una argumentación acerca de por qué algunas de estas categorías, aunque sean construcciones sociales, se toman en cuenta al estudiar fenómenos de contacto lingüístico (p. 230).

De otro lado, al mirar de cerca los temas desarrollados por capítulo, encontramos que las autoras procuraron cubrir los principales aspectos de un enfoque sociopolítico del español en los Estados Unidos, para lo cual dedican espacio a contenidos focalizados en cuestiones históricas, raciales, políticas, entre otros. Como ya se comentó en la sección 1, el primer capítulo sirve de introducción al establecer el marco conceptual y la organización de todo el libro. Luego, el capítulo 2, “The Demographics of Spanish in the US”, desarrolla un panorama cuantitativo general sobre los hablantes de español en el país basado sobre todo en estadísticas censales. Estos datos se presentan y discuten tomando en cuenta criterios relevantes que caracterizan a los hablantes, como su lugar de origen y otros rasgos que muestran su diversidad, y su habilidad y uso de la lengua. La segunda parte del capítulo pone atención en los efectos vinculados a estos últimos criterios, como el mantenimiento y la pérdida de la lengua en favor del inglés.

El capítulo 3, “The History of Spanish and Spanish-speakers in the US”, realiza un recuento histórico de la presencia del español y de sus hablantes en los Estados Unidos, empezando por las primeras expediciones y asentamientos españoles en territorios norteamericanos. Después, se centra en las fases de expansión de los Estados Unidos ya como país independiente, las que, en cuanto a su relación con el español y la población hispanohablante, corresponden a la incorporación de la Florida como primera fase, el amplio suroeste anexado después de la guerra mexicano-estadounidense como segunda y Puerto Rico como tercera. La última sección del capítulo se dedica a la situación más pertinente para la época actual, es decir, la inmigración desde países hispanohablantes.

El capítulo 4 se titula “Language Ideologies” y es con seguridad uno de los más teóricos de toda la obra. Fuller y Leeman extienden el marco conceptual en el que se basan presentando la noción de ideologías lingüísticas, su rol intermediario entre lo lingüístico y lo social, y las consecuencias que pueden producir en todo nivel de la sociedad. Para ello, explican las ideologías lingüísticas más estudiadas y que mejor se adecuan al contexto del español en los Estados Unidos: la ideología de la lengua estándar, la ideología de una nación-una lengua, la ideología de suma cero (*zero-sum ideology*) y el monolingüismo normativo, entre otras. Una observación particular que se puede hacer a este capítulo es que en el desarrollo de la teoría sobre ideologías lingüísticas, las autoras se basan, entre otros, en Kroskrity (2004). Este autor plantea cinco dimensiones para abordar el análisis de las ideologías lingüísticas; de estas cinco, al menos cuatro se pueden distinguir a lo largo del capítulo de Fuller y Leeman, aunque no se precisa de forma evidente que sean tales dimensiones. Sin embargo, la quinta, que sostiene que las ideologías lingüísticas se usan productivamente para crear y representar identidades sociales y culturales diversas, no se menciona explícitamente a pesar de que hay ejemplos propuestos (p. 65) y algunas alusiones indirectas a ella (p. 68). El que este capítulo dedicara un espacio saliente a esta quinta dimensión hubiera sido fundamental para establecer una conexión más evidente con el capítulo 6, cuyo tema central es la relación del lenguaje con la identidad.

El capítulo 5 se llama “Race, Racialization and Latinx Ethnoracial Identity”; como dice su nombre, aborda el estudio de la raza como construcción social y de la racialización

como la extensión de los significados raciales hacia otros aspectos sociales con los que no habían tenido alguna asociación previa (p. 94). Así, se parte de una discusión principalmente teórica, en la que también se desarrolla el concepto de etnicidad, para luego enfocarse en cómo estas nociones son relevantes para el caso de América Latina y luego para los Estados Unidos. En este último contexto, el capítulo destina su sección final al análisis de los censos, que brindan información valiosa sobre las categorías usadas para identificar grupos. De este modo, queda claro que este capítulo es medular en el libro, porque desarrolla conceptos teóricos que servirán como base para otros temas tratados en los siguientes apartados. Lo que extraña, no obstante, es que se hayan incluido secciones bastante breves para hablar de elementos más centrados en lo lingüístico, como la que se refiere a la racialización del español o la que complementa las categorías censales con la discusión sobre el lenguaje inclusivo. La primera, por ejemplo, permite que se anticipen o se recuerden problemáticas citadas en otros capítulos, pero solo de una manera muy sintetizada, como para que la explicación más profunda se busque en esas otras páginas. Esta observación deja la sensación de que no era el objetivo central del capítulo el detenerse en los aspectos más lingüísticos, por lo que habría sido recomendable elaborar el tema de raza y racialización desde una perspectiva social y política, y más bien aplicar ese marco teórico a temas lingüísticos a lo largo de los otros capítulos. De hecho, referencias fundamentales, como el análisis de Rosa y Flores (2017) acerca de cinco componentes claves para entender la perspectiva raciolingüística (y aplicarla al caso del español en los Estados Unidos), se mencionan brevemente en esa sección. Así, se desaprovecha la oportunidad de elaborar una argumentación teórica mucho más rica que podría haber encajado bien en este capítulo o incluso en otros en los que las temáticas lingüísticas son el eje central.

El mayor énfasis en el vínculo entre lengua e identidad se encuentra en el capítulo 6, “Language and Identity”. Tal como ocurre con el cuarto capítulo, este también elabora propuestas notoriamente teóricas, tales como el estudio de la identidad y la construcción de esta a través de la interacción. Para ello, se parte del trabajo de Bucholtz y Hall (2005), en el que la identidad es vista como el producto que se va construyendo a través de nuestro comportamiento lingüístico y social. No es una sola, sino que pueden emerger de distintas maneras; incluso se pueden analizar como identidades múltiples e interseccionales. El capítulo explica al detalle los cinco principios de Bucholtz y Hall acerca de la construcción de identidades en interacción y concluye con una revisión sobre el estudio del *mock Spanish*, que hace referencia a discursos racializadores en que se dicen frases que se asemejan al español, pero que realmente no constituyen un intento de hablar la lengua.

El capítulo 7, “Spanish and Spanish-speakers in US Media”, se centra en la presencia del español y de personas latinas en los medios de comunicación, principalmente en el cine y la televisión. De esta manera, se hace un repaso de numerosos estudios de caso que analizan los estereotipos recurrentes, la participación de latinxs en programas de lengua inglesa y en programas de canales hispanohablantes. También se incluye una sección dedicada a la representación de la población latina en las noticias. Por último, el capítulo se enfoca en la presencia del español en el paisaje lingüístico, es decir, el uso de la lengua en ambientes públicos por medio de carteles, avisos, grafiti, etc. Este último punto queda parcialmente desconectado de las secciones anteriores, ya que no se hace muy evidente cómo se engarza con el tema del español en los medios de comunicación. Por otro lado, un comentario general sobre este capítulo es la justificación de que sea

efectivamente un capítulo completo. Si bien resulta de interés conocer la gran cantidad de investigaciones que se han realizado acerca de series de televisión, la información nueva o la discusión teórica que se aporta es menor y podría haberse tratado como parte de un marco teórico más general. Por ejemplo, muchos estudios mencionados contribuyen significativamente al argumento de que el español es una lengua racializada en los Estados Unidos, de forma que podrían guardar mucha más coherencia si fueran parte del capítulo 5 sobre racialización. En el mismo sentido, otros estudios citados hacen referencia a la construcción de la identidad de los latinxs, así que serían útiles para ejemplificar nociones del capítulo 6.

En cuanto al capítulo 8, “Language Policy and Spanish in the US”, se parte de una explicación teórica sobre planificación lingüística, las medidas que involucra y las orientaciones que se toman con respecto a las lenguas, para luego pasar por una revisión histórica de las políticas aplicadas en el país y la polémica acerca de tomar al inglés como lengua oficial. Este capítulo se conecta bastante bien con el siguiente, “Spanish in US Schools”, en el que se describen y evalúan los tipos de programas implementados para hablantes de otras lenguas, sobre todo actualmente, y las medidas consideradas para grupos específicos, como los hablantes de herencia. La segunda mitad del capítulo 9, por otra parte, aplica la teoría de ideologías lingüísticas vista en el capítulo 4 a los casos educativos y termina con una visión reciente desarrollada a partir de la pedagogía crítica. Así, un comentario general con respecto a los capítulos 8 y 9 es que bien podrían haberse organizado en uno solo, debido a que sus contenidos se complementan adecuadamente, además de que el segundo se vale también de lo que ya se ha presentado en capítulos anteriores, como el de ideologías lingüísticas.

El capítulo 10, “Structural Aspects of Speaking Spanish in the US”, es el que más se distancia de los demás, puesto que trata sobre aspectos formales en su mayor parte: variedades lingüísticas del español de los Estados Unidos, fenómenos de contacto lingüístico como el *codeswitching* o cambio de código, el contacto de variedades de español en grupos inmigrantes y las variedades de inglés habladas por latinxs. Además de la descripción y explicación de rasgos léxicos, fonológicos y gramaticales, el capítulo incluye una sección final que regresa hacia el lado social al presentar estudios sobre actitudes lingüísticas hacia las variedades de español y hacia el bilingüismo. Esta intención de hacer explícito el vínculo con la parte social incluso en un capítulo que habla sobre estructuras lingüísticas resulta acertada, pues así no se percibe como una desviación del tema central del libro. Sin embargo, esto podría haberse aprovechado un poco más en las secciones descriptivas sobre las variedades, como al hablar de variación léxica, fonética y estructural, o en la sección en que se presenta el caso de los pronombres. Es posible agregar información, al menos general, con respecto a la valoración o la percepción sociolingüística que tienen los hablantes acerca de estos rasgos.

El último capítulo del libro se titula “The Future of Spanish in the US” y cumple dos propósitos al mismo tiempo: incidir en los temas fundamentales vistos en los capítulos anteriores y plantear posibles escenarios futuros para el español en los Estados Unidos. Por este motivo, está organizado en secciones que corresponden a cada uno de los capítulos previos. Esta forma de organización no necesariamente es adecuada para elaborar pronósticos, puesto que si bien hay mucho material para tomar en cuenta acerca de cómo se vislumbran las direcciones que tomará el español con respecto a las temáticas de algunos capítulos anteriores, en otros más bien la pregunta sobre lo que se espera en

un futuro parece quedar abierta y, por tanto, no hay muchos detalles para elaborar. Por el contrario, la última sección, sobre conclusiones y conexiones, podría haber sido más amplia al tratarse de la sección que cierra todo el libro. Una pregunta grande que surge de la lectura de los capítulos es que hablar español en los Estados Unidos tiene muchos y diversos significados teniendo en consideración quiénes lo hablan, por qué lo hablan y en qué contextos. Una discusión sobre cómo responder a una pregunta general de este estilo podría haber sido provechosa para esa última sección. También, las autoras incluyen en ese mismo apartado sugerencias sobre qué puede hacer el público lector desde su posición en favor del español y de la población latina. Estas sugerencias constituyen un punto a favor del capítulo de cierre, pues conectan a quien lee directamente con los temas tratados y le crean el compromiso de participar activamente. Por esa misma razón, se podría considerar una mayor y más detallada elaboración, es decir, que no se trate solo de un párrafo, sino de una sección que le dé mayores herramientas a quien decide involucrarse.

5. Apreciación final

El libro es una publicación de gran relevancia para quienes estudian o tienen interés en conocer más sobre el español en los Estados Unidos, sobre todo si buscan entender la situación de la lengua y de sus hablantes a partir de contextos más amplios que incluyan aspectos políticos, sociales e ideológicos. Por ello, es altamente recomendable para estudiantes o para investigadores que precisen de referencias sobre temas particulares concernientes al español. La organización del libro y de cada capítulo ayuda de forma notoria a que una persona que no necesariamente cuenta con una formación lingüística pueda seguir la argumentación e ir más allá al aplicar lo aprendido a casos distintos a través de las actividades y preguntas de discusión. Por otro lado, hemos precisado que hay capítulos que desarrollan contenidos muy claros y bien sustentados, pero también hay otros en que algunas secciones adolecen de cohesión con el resto del capítulo o no parecen ofrecer un contenido temático que amerite desarrollarse como un capítulo independiente. Asimismo, con el objetivo de que sea un libro bastante manejable incluso para estudiantes, observamos que las referencias que se utilizan pueden organizarse de manera más efectiva. Aparte de estos últimos comentarios, creemos que el libro representa un aporte valioso para los estudios sociolingüísticos.

Referencias bibliográficas

Bucholtz, Mary y Kira Hall. 2005. Identity and interaction: a sociocultural linguistic approach. *Discourse Studies* 7. 4/5: 585-614.

Burr, Vivien. 2015. *Social Constructionism*. 3ª edición. East Sussex/Nueva York: Routledge.

King, Sharese. 2020. From African American Vernacular English to African American Language: Rethinking the Study of Race and Language in African Americans' Speech. *Annual Review of Linguistics* 6: 285-300.

Kroskrity, Paul. 2004. Language Ideologies. En A. Duranti, ed. *A Companion to Linguistic Anthropology*. Nueva Jersey: Wiley-Blackwell, pp. 496-517.

Lanehart, Sonja y Ayesha M. Malik. 2015. Language Use in African American Communities: An Introduction. En S. Lanehart, ed. *The Oxford Handbook of African American Language*. Oxford: Oxford University Press, pp. 1-19.

Rosa, Jonathan y Nelson Flores. 2017. Unsettling race and language: Toward a raciolinguistic perspective. *Language in Society* 46.5: 1-27.

Notas

¹ Fuller y Leeman aclaran en el capítulo 1 que usarán el término “latinx” como genérico inclusivo antes que otras opciones como “latino”, “latina/o”, “latin@”. Mayor discusión sobre el uso del lenguaje inclusivo se desarrolla en el capítulo correspondiente a racialización e identidad.

² Este es el nombre que King (2020) usa a lo largo de su artículo aduciendo, a partir de Lanehart y Malik (2015), que se trata de un término neutro y que involucra todas las variedades de las comunidades afroamericanas (p. 287). King presenta la referencia como Lanehart (2015) para aludir a la edición completa, aunque la referencia exacta proviene del capítulo introductorio, escrito por Lanehart y Malik.

Emma Gallardo Richards (Universidad Autónoma de Barcelona). Reseña de Dubert-García, Francisco; Míguez, Víctor; Sousa, Xulio. 2020. *Variedades lingüísticas en contacto na Península Ibérica*. Santiago de Compostela (España): Consello da Cultura Galega

El libro *Variedades lingüísticas en contacto na Península Ibérica*, editado por Francisco Dubert-García, Víctor Míguez y Xulio Sousa, de la Universidade de Santiago de Compostela, y publicado en 2020 por el Consello da Cultura Galega, reúne siete trabajos que abordan el contacto de distintas lenguas y variedades lingüísticas en el ámbito peninsular. La obra parte de una visión del contacto lingüístico como un hecho genuino y consustancial a todas las lenguas, tanto sincrónica como diacrónicamente. Los trabajos recogidos en esta obra, a sabiendas de la pluralidad de enfoques desde los que se puede tratar este ámbito de investigación, se aproximan al objeto de estudio desde una visión lingüística –“a perspectiva que se pretende adoptar ten en conta os fenómenos estritamente lingüísticos que o contacto produce” (p. 8)– y atendiendo a las distintas variedades que se restringen geográficamente a la península Ibérica. Así, junto con el contacto del castellano con el catalán (capítulo 1), con el asturiano (capítulo 2), con el vasco (capítulo 4) y con el gallego (capítulos 6 y 7), se insertan estudios que plantean las relaciones fronterizas entre el mirandés y el portugués (capítulo 3), así como entre las variedades meridionales del andaluz (capítulo 5). De este modo, se logra dar cuenta de una parte significativa del escenario lingüístico peninsular que no siempre ha gozado del mismo estatus en las investigaciones científicas. Además, hay que resaltar que la obra engloba artículos de índole teórica, metodológica y experimental, que al mismo tiempo prestan atención a distintos planos de la lengua.

En las páginas introductorias de la obra, “Limiar” (pp. 6-12), se presentan las bases del contacto lingüístico para situar al lector, tanto a un público especializado o a quien no posee un bagaje de conocimientos previo. Para ello, a través de ejemplos extraídos principalmente del contexto gallego, no solo se hace hincapié en las posibilidades de estudio del contacto lingüístico, sino que también se reflexiona acerca de cómo se produce, los factores y mecanismos que intervienen, su vínculo con el cambio lingüístico y la percepción social negativa que para algunos hablantes tiene debido a motivos identitarios. Asimismo, como ya se ha hecho referencia, se clarifica la perspectiva de estudio adoptada y se expone de forma somera las principales contribuciones de cada capítulo.

El capítulo que inaugura la obra, “Opinión, medición y generalización como retos de la lingüística de contacto. Ejemplos del estudio del contacto entre castellano y catalán” (pp. 17-46), corre a cargo de Carsten Sinner. En estas páginas se lleva a cabo un verdadero ejercicio crítico y de revisión sobre las distintas creencias, generalizaciones e imprecisiones metodológicas que han impregnado ciertos estudios publicados desde la década de 1970 sobre el contacto entre estas dos lenguas y que se deben, en gran parte, a errores de base empírica que desfiguran los resultados y las interpretaciones. Sin embargo, acerca de los trabajos pioneros en esta materia se afirma que “no hay que dejar de subrayar que constituyen loables intentos que no deben descontextualizarse ni criticarse de forma anacrónica” (p. 23), de modo que el foco se traslada a aquellas

investigaciones posteriores que se sostienen sobre ellos y que, por ende, no han reparado en el diseño y en los procedimientos empleados para la recogida de datos. Son dos las ideas sobre las que con más fuerza se incide y se vertebra el capítulo: por una parte, la direccionalidad del contacto lingüístico, ya que se establece que puede producirse en ambos sentidos y, por otra, el grado del dominio de la lengua materna sobre otra con la que se produce el contacto. En relación con estas cuestiones, se pondera la noción de *interferencia* y sus vínculos con la *integración*, se pone en valor la importancia de tener en cuenta las variedades del catalán y del castellano y se reflexiona sobre la fiabilidad de los perfiles lingüísticos a la hora de evaluar el grado de bilingüismo de un informante. En definitiva, la aportación de Sinner constituye una lectura imprescindible no solo para quienes investigan el contacto entre el catalán y el castellano, sino también para aquellos que se centran en el examen de situaciones de contacto con otras lenguas y variedades, puesto que las advertencias metodológicas que se ofrecen son, en realidad, intrínsecas a cualquier investigación que siga el método científico.

Ramón de Andrés es el autor del segundo capítulo, “Contacto de lenguas en Asturias: asturiano, «amestáu» y castellano” (pp. 47-79), en el que se presenta una panorámica sobre los fenómenos lingüísticos que puede ocasionar el contacto entre el castellano y el asturiano, para los que resalta que se “debe tener en cuenta el sistema dialectal de la base contrastiva, que puede diferir de unas hablas a otras” (p. 57). Tras repasar los principales fenómenos de interferencia y de alternancia de código entre ambas lenguas, se centra la atención en el *amestáu*, esto es, el asturiano castellanizado, que resulta de una hibridación lingüística desigual y en cuya mezcla se aprecia un mayor dominio del castellano sobre el asturiano. El autor da cuenta de sus posibilidades de estudio, que pasan por enfoques glotológicos o perceptivos, según si se pretende ahondar en cuestiones puramente lingüísticas o bien ideológicas. Para ello, se sirve, por ejemplo, de un estudio anterior (Andrés 2017) para demostrar cómo puede tratarse estadísticamente el índice de *asturianidad* o de *castellanidad* de una muestra. Al mismo tiempo, se vale de encuestas sociolingüísticas procedentes de otras investigaciones llevadas a cabo en el ámbito asturiano –siempre citadas, pero a las que el lector debe acudir para comprobar el índice de representatividad– para extraer datos que permiten arrojar cierta luz al *amestáu*, ya que, como el propio autor afirma “no ha venido siendo asunto prioritario de estudio” (p. 70). En suma, las directrices de estudio propuestas por Andrés no solo son válidas para *amestáu*, sino también son extrapolables a otros trabajos sobre la hibridación lingüística.

En el tercer capítulo, que lleva por título “Consecuencias do contacto lingüístico entre mirandés e portugués” (pp. 81-100), Alberto Gómez Bautista se interesa por los fenómenos de contacto lingüístico entre el portugués y el mirandés y aporta una taxonomía de casos divididos según su pertenencia al plano ortográfico, fonético-fonológico, morfosintáctico y léxico. A la hora de contextualizar los elementos analizados, se llama la atención sobre dos puntos: en primer lugar, se matiza que, debido al contexto sociolingüístico y a que el mirandés carece de una variedad normativa, el portugués ejerce un influjo mayor sobre él y, en segundo lugar, se incide en la pertinencia de tener en cuenta la variedad concreta del portugués que influye. Igualmente, en ciertos fenómenos, se manifiesta la imposibilidad de determinar si realmente se trata de una influencia del portugués o de una evolución propia del mirandés. Así, aunque esta contribución ofrece una caracterización predominantemente

fonético-fonológica y morfosintáctica de las relaciones entre el portugués y el mirandés, el autor reconoce que “ten a ver co feito de ser o campo máis estudado por nós e non é debido necesariamente a unha maior exposición desta parte da gramática da lingua á influencia do portugués” (p. 98).

Sara Gómez Seibane se ocupa de la cuarta aportación, “Patrones de convergencia en lenguas tipológicamente no relacionadas: lengua vasca y castellano” (pp.101-126). La autora, tras resumir los principales rasgos fonológicos, morfológicos y sintácticos que caracterizan al euskera y al castellano, aporta las claves históricas que marcaron el contacto lingüístico entre ambas lenguas. En este sentido, se destaca que el contacto únicamente se había producido, hasta el siglo XIX, en las élites y que, por tanto, el castellano había permanecido ajeno a la mayor parte de la población. Así, Gómez incide en la centuria mencionada como el momento en que la influencia entre las dos lenguas fue mayor debido al incremento de la escolarización, a raíz de que en el sistema educativo era reglamentario el uso del castellano (*cf.* Fernández de Gobeo 2018). En consecuencia, un número significativo de hablantes pertenecientes a las clases populares, que hasta el momento habían sido monolingües en euskera, empezaron a ser conocedores y a emplear la lengua nacional, la española. También se aborda este contacto desde un punto de vista lingüístico, deteniéndose en los fenómenos que podrían aducirse a la relación entre las dos lenguas. El último bloque del capítulo se centra en la convergencia lingüística entendida como motor que favorece la modificación estructural de una lengua meta por influencia de otra, impulsada por una “replicación de patrones” (Matras 2009: 238), que puede producirse incluso cuando intervienen sistemas lingüísticos tipológicamente divergentes, como es el caso del vasco y del castellano. Por ello, la autora alude y explica algunos de los fenómenos de convergencia en el castellano del País Vasco, así como en la lengua vasca. De esta manera, el lector encontrará en esta contribución un panorama indispensable sobre los efectos del contacto lingüístico entre el vasco y el castellano, especialmente en lo relativo al fenómeno de la convergencia.

El foco del quinto capítulo, “Procesos de estandarización y divergencia dialectal en las variedades meridionales del español de España” (pp. 127-154), elaborado por Matilde Vida-Castro, son las variedades del español meridional, entre las que se destaca el español de Málaga, que ha sido abordado en profundidad por el grupo de investigación Vernáculo Urbano Malagueño (VUM) de la Universidad de Málaga, al que pertenece la autora. Concretamente, se examinan desde un punto de vista de la pronunciación dos patrones que no son exclusivos del panorama lingüístico andaluz, sino que también se observan en otros escenarios lingüísticos en el resto de Europa. Por una parte, se encuentran los procesos de convergencia del sociolecto de origen hacia el estándar, que al mismo tiempo están ligados al surgimiento de una variedad intermedia coinética, fruto de una nivelación, con una mayor incidencia en las áreas urbanas. Esta última combina rasgos meridionales con los del español estándar, cuya aparición viene determinada –aunque no siempre, como sucede con el seseo– en la mayoría de los casos por el nivel de instrucción, como se ha comprobado desde el grupo VUM. Por otra parte, está la divergencia dialectal, mediante la que los hablantes bien se alejan del estándar español o de las variedades andaluzas de prestigio. El capítulo da cuenta, en el quinto epígrafe, de las motivaciones que pueden o no inclinar a los hablantes a la estabilidad dialectal, esto es, al mantenimiento de las características lingüísticas regionales, decisión que parece estar regida por dos variables: dónde surge la variedad y

cómo se conciben los rasgos por parte de los hablantes. Esto se ejemplifica por medio de estudios realizados en las últimas décadas que apuntan a tendencias sociales generales sobre el mantenimiento de rasgos dialectales propios y que se completan con la alusión a algunos trabajos que se han ocupado de casos particulares y grupos sociales más reducidos. El estudio de Vida-Castro invita, sin duda, a seguir ahondando en las causas y en las distintas ideologías que se encuentran detrás de las decisiones lingüísticas de los hablantes andaluces ante el prestigio del estándar español o del regional.

El sexto capítulo constituye la primera de las dos contribuciones que atiende a cuestiones relativas al contacto lingüístico con el gallego. Corre a cargo de Alba Agüete Cajiao y lleva por título “Contacto de lenguas e variedades na emerxencia de novos modelos de vocalismo galego” (pp. 155-194). En este estudio, enmarcado en la fonética acústica, Agüete analiza los fenómenos de variación en el vocalismo gallego en tres posiciones: tónica, átona a inicio de palabra y átona a final de palabra. La importancia de su investigación radica en que se incide en las consecuencias que puede desencadenar en el gallego su contacto con el castellano, además de la procedencia geográfica. Esta cuestión es relevante por las diferencias entre el sistema vocálico gallego y castellano, compuestos por siete y cinco fonemas respectivamente –el gallego distingue entre /e/–/ɛ/ y /o/–/ɔ/, contraste ausente en castellano–. Para la obtención de los datos, la autora recurrió a la grabación de cincuenta frases con sentido completo por parte de doce universitarios de diferentes puntos de Galicia cuya lengua materna y habitual era el gallego. Los resultados generales de la muestra analizada indican la existencia para el vocalismo tónico y pretónico a inicio de palabra de dos tendencias en los hablantes gallegos: por una parte, el mantenimiento de la oposición /e/–/ɛ/ y /o/–/ɔ/ exclusiva en algunos de los informantes de procedencia rural. Por otra parte, la autora revela otro modelo en el que se pierde dicho contraste tanto por informantes rurales como urbanos y en el que, por tanto, parece reproducirse el paradigma vocálico castellano formado por cinco fonemas. Por el contrario, los resultados que arroja el vocalismo átono a final de palabras son de una cierta homogeneidad y en los que el castellano parece no repercutir. Por tanto, el trabajo de Agüete revela la influencia del castellano en el vocalismo gallego, que como consecuencia constituye un modelo innovador caracterizado por suprimir la distinción entre /e/–/ɛ/ y /o/–/ɔ/ a favor de /e/ y de /o/ y, así, regirse por el sistema vocálico del castellano. A su vez, alude a la posibilidad de que la pérdida del contraste se deba a una evolución interna de la lengua por un proceso de fusión presente en otras lenguas románicas. En suma, los resultados obtenidos de esta investigación, pese a reducirse a una muestra de doce informantes, patentizan la necesidad de seguir explorando las consecuencias del contacto lingüístico entre el gallego y el castellano mediante las técnicas experimentales de la fonética acústica, dado que es un campo que, como recuerda Agüete, se ha tratado en pocas investigaciones científicas hasta el momento.

El séptimo y último capítulo del libro se titula “*Había* + participio no español falado en Galicia: un estudo de corpus” (pp. 195-220) y está elaborado por Victoria Vázquez Rozas. Por segunda vez en la obra, se incluye una contribución referente al gallego. En este caso, se toma en cuenta la forma verbal compuesta *había* + *participio* en el castellano de Galicia a partir de los datos obtenidos de 53 entrevistas semidirigidas que provienen del corpus oral ESLORA. Esta cuestión se vuelve significativa al tomar en cuenta los trabajos que señalan la tendencia en el castellano de Galicia a la

simplificación de las formas verbales compuestas, como sucede con la distinción entre el indefinido y el pretérito perfecto. En cambio, la aportación de Vázquez supone una aportación tanto al reparto y al uso de una forma compuesta, *había + participio*, que, por influencia del gallego, tiene valores distintos al castellano: además del antepretérito o pluscuamperfecto (una acción anterior a otra anterior), puede expresar pretérito y copretérito. Por esto, la autora también apunta a confluencia con las formas *cantara* – con la que comparte los valores en indicativo– y *cantase* –esta última forma por el valor subjuntivo que también posee *cantara*–, a las que remite igualmente en su análisis. La estratificación de los datos del corpus en tres grupos sociolectales que tienen en cuenta el nivel de formación de los informantes –estudios primarios, medios o universitarios– permiten a Vázquez extraer algunas conclusiones sobre la distribución de las formas verbales. Se apunta a la vitalidad de *había + participio* en el castellano de Galicia para la expresión de los tres valores ya mencionados, se arguye el papel de la enseñanza, que no legitima *cantara* con valor indicativo, como motor de una ultracorrección y uso de la estructura compuesta en la expresión de pretérito y copretérito y, por último, se aclara el reparto de las formas de subjuntivo en *-ra* y en *-se* basado en distinciones sociolectales, vinculadas de nuevo a evitar la primera de las formas cuanto mayor es el grado de instrucción por desautorización del alejamiento de la norma estándar. Como reconoce Vázquez en las conclusiones, el estudio debería completarse en un futuro con información sociolingüística y metalingüística de cada informante, para así poder arrojar más luz sobre este aspecto de la morfología verbal del castellano de Galicia.

En definitiva, el libro *Variedades lingüísticas en contacto na Península Ibérica* tiene la virtud de agrupar diversas muestras del contacto entre distintas lenguas y variedades de un espacio geográfico delimitado, permitiendo así al lector conocer las diferentes posibilidades de estudio en clave variacional que presenta el panorama lingüístico peninsular. De la misma forma, es destacable la atención a los fenómenos en diferentes planos de la lengua, las aproximaciones mixtas –teóricas, metodológicas y experimentales– a los objetos de estudio, así como la proyección del provecho de distintos corpus, proyectos, archivos y herramientas para las investigaciones de este ámbito.

Referencias bibliográficas

- Andrés Díaz, Ramón de. 2017. Índiz d'asturianidá de dos testos falaos: un ensayu. *Archivum* 67: 41-88.
- Fernández de Gobeo Díaz de Durana, Nerea. 2018. *La enseñanza del castellano en las escuelas vascas del siglo XIX: Estudio histórico y lingüístico*. Bilbao: Servicio Editorial de la Universidad del País Vasco.
- Matras, Yaron. 2010. Contact, Convergence, and Typology. En Raymond Hickey, ed. *The Handbook of Language Contact*. Malden: Wiley-Blackwell, pp. 66-85.

Julieta Del Prato (Universidad Nacional de la Patagonia San Juan Bosco). Reseña de Carranza, Isolda E. 2020. *Narrativas interaccionales. Una mirada sociolingüística a la actividad de narrar en encuentros sociales*. Córdoba (Argentina): Editorial de la Facultad de Lenguas.

Narrativas interaccionales. Una mirada sociolingüística a la actividad de narrar en encuentros sociales de Isolda E. Carranza se destaca por reunir, con un enfoque sociolingüístico interaccional, estudios sobre el narrar (y su combinación con la actividad de argumentar) considerado en encuentros sociales de diverso tipo, incluyendo los que se realizan como parte de prácticas socioculturales establecidas. El eje vertebrador que organiza los trabajos es la narración y la interacción cara a cara como espacio de reproducción o transformación de aspectos de estructuras sociales. Al vislumbrarse allí el juego de expectativas, visiones de mundo y posturas asociado a los roles discursivos y situacionales de los participantes, se vuelve fundamental –en el plano lingüístico discursivo– el estudio de las combinaciones entre los modos narrativo y argumentativo del discurso.

Este libro propone un planteo integral para el abordaje tanto de la actividad de narrar en diferentes contextos sociales, como de los productos textuales que son resultado de esas prácticas. Los análisis recuperan diversos géneros discursivos orales y escritos, como conversaciones espontáneas, entrevistas etnográficas, alegatos penales, declaraciones testimoniales y actas oficiales del ámbito judicial. La propuesta global de la obra responde a tres supuestos básicos: a) la relación entre el mundo diegético y el mundo de la interacción en el que emerge la práctica social de narrar, condicionada por un micro y un macrocontexto; b) la multilateralidad de la producción narrativa, resultado de una negociación interaccional entre los participantes del evento comunicativo; y c) la necesaria reflexión de la investigadora respecto de su rol en la recolección y construcción de datos.

La perspectiva de la sociolingüística interaccional adoptada contempla la dinámica de las relaciones de poder entre los participantes, las negociaciones por la representación del mundo que allí se producen, la construcción de la voz identitaria que cada sujeto asume en el seno de una interacción y la dimensión ideológica. A diferencia de las corrientes pragmáticas que se concentran en contextos situacionales acotados, Carranza plantea un modelo de análisis multinivel que abarca un nivel lingüístico y retórico, un nivel intermedio de práctica social, un nivel de contexto macrosocial y las relaciones de retroalimentación que entre ellos se generan. Las explicaciones de los fenómenos observados son el resultado de la aplicación de conceptos vigentes en los estudios narrativos y de la argumentación –como postura epistémica, cronotopo o evidencia– y son enriquecidas con novedosas herramientas conceptuales, entre las que aquí mencionamos, tesis del relato, plausibilidad narrativa, narrativas de baja narratividad, temporalidad debilitada, macro relato y autoridad retórica.

En cuanto a los aspectos metodológicos, todos los estudios incluidos en este libro surgen a partir de comprometidas etnografías que propician un conocimiento profundo de los contextos socioculturales puestos en foco. Frente a la base empírica, la autora

expone un reflexivo y necesario grado de vigilancia epistemológica en todas las etapas de su práctica de investigación. Cabe destacar la profusa bibliografía integrada críticamente y como resultado de un amplísimo relevamiento que incluye referencias ineludibles y publicaciones del mismo año en que esta obra salió a la luz.

Respecto de los destinatarios, el libro está orientado principalmente a investigadores y estudiantes de posgrado del área de las ciencias del lenguaje y la comunicación, que deseen profundizar en el estudio de textos narrativo-argumentativos interaccionales, trascendiendo el tratamiento localizado en la micro situación comunicativa. Por esto, también es una herramienta primordial para investigadores de aquellas ciencias sociales que le otorgan relevancia al uso del lenguaje y a los discursos. Cabe destacar que el hecho de que esta obra, de tal envergadura, haya sido publicada en Argentina, en español y con condiciones de publicación de acceso abierto (en el repositorio digital de la Universidad Nacional de Córdoba, Argentina) resalta su contribución al panorama académico latinoamericano y le otorga potencial para estimular el quehacer científico de la región.

Este volumen consta de catorce capítulos distribuidos en cinco partes. Dos recursos son de suma utilidad para orientar la lectura: cada una de las partes contiene una introducción a los capítulos que la componen y, al final de cada capítulo, se ofrece una síntesis del argumento central desarrollado. Estas herramientas posibilitan la lectura de los capítulos de manera separada, ya que permiten contextualizar rápidamente el lugar que ocupa en el mapa global del libro.

La Parte I, “Perspectiva interaccional, el concepto de voz y la explotación de recursos formales”, incluye los capítulos 1, 2 y 3: allí, además de la historización del estado de la cuestión, el énfasis está puesto en la construcción dialógica de la figura del narrador y de las estrategias retóricas y de *performance* implementadas durante la interacción.

En el capítulo 1, “Confluencias teóricas en el estudio de la narratividad interaccional”, la autora recupera los principales desarrollos de los estudios interaccionales de la narración: las contribuciones de la etnometodología, del interaccionismo simbólico y de reconocidos sociolingüistas, quienes han explorado la vinculación entre estas narrativas y la construcción de un *self*, valorando aquellas investigaciones que conectan lo personal, lo interpersonal y lo macrosocial. Este capítulo introductorio finaliza con algunas precisiones terminológicas y otras referidas a las pautas de transcripción. En el capítulo 2, “Dialogismo y autoridad retórica”, se profundiza en el carácter dialógico de las conversaciones espontáneas. Se estudia la construcción de “autoridad retórica” como una estrategia de configuración de la imagen social del yo. El efecto de autoridad se logra explotando dos recursos: la reducción de la distancia intertextual con las voces citadas y el discurso razonado. En el capítulo 3, “*Performance*: la forma puesta en primer plano”, el énfasis está puesto en la función persuasiva que cumplen los rasgos formales de la “actuación”, como el paralelismo y la prosodia. En el género “entrevista”, se estudian los recursos mediante los cuales se crea el efecto de verosimilitud. En las narrativas de acusados en testimonios judiciales, se demuestra que las formas de discurso referido se distribuyen, de modo sostenido y estratégico, entre los personajes de los relatos.

En la Parte II, “Narrar para la audiencia: Tesis y Tiempo”, que contiene los capítulos 4 y 5, la autora indaga en los mecanismos de construcción y justificación de sentidos asociados a experiencias de vida de inmigrantes latinoamericanos.

El capítulo 4, “Los relatos en entrevistas y sus tesis”, estudia cómo se construye la “tesis del relato” en el discurso coproducido por el par entrevistado-entrevistador. La tesis del relato es definida como aquella proposición que constituye una declaración controversial para el contexto de enunciación, y que es apoyada por elementos narrativos. Cuando la tesis está explícita, puede preceder a la narración que la legitima o puede aparecer luego de la narración, indicando cómo debe interpretarse la experiencia. Cuando la tesis está implícita, el narrador la considera potencialmente refutable y ello justifica la aparición de la narración. En el capítulo 5, “Los acontecimientos hipotéticos y los repetidos o habituales”, interesada en variedades de narrativas que no constituyen relatos, Isolda Carranza acuña el concepto de “baja narratividad” para abarcar las acciones del pasado que no pueden ordenarse linealmente y las que no son discretas (de aspecto perfectivo). Por un lado, muestra que la narración de acontecimientos contrafactuales puede adquirir una función evaluativa, focalizadora o explicativa respecto del relato abarcador. Por el otro, estudia las acciones reiteradas o habituales en el pasado, que recrean la visión que el sujeto tiene del telón de fondo donde se sitúan los hechos focalizados, por lo que también se convierten en un aporte para el análisis ideológico.

La tercera parte del libro, “El mundo de la narración y los límites permeables del mundo diegético” (capítulos 6, 7 y 8), hace hincapié en el activo rol de la audiencia en el desarrollo de la narrativa y en los modos de intersección entre los planos intra y extradiegéticos. Los tres capítulos se desarrollan en un contexto de resistencia al prejuicio étnico generalizado.

El capítulo 6, “La réplica fuera del mundo diegético”, se ocupa de la interfaz que se produce cuando el relato representa una confrontación entre un personaje-oponente y un narrador-proponente y se estudia la proyección de roles actanciales de la diégesis sobre los roles situacionales en la entrevista. En el capítulo 7, “El personaje a cargo de la lógica retórica”, el centro del análisis son dos movimientos retóricos, la analogía y la explicación (causal), utilizados en una confrontación argumentativa en el mundo diegético. Como una marca característica de este libro, la autora cierra el capítulo con una reflexión que atañe a la labor de los investigadores: el análisis lingüístico y discursivo debe ser completado con la consideración de las condiciones sociales estructurales en las que se da el encuentro con los sujetos investigados. En el capítulo 8, “Razonamiento, el sí mismo y el otro”, Carranza aplica el concepto de “postura” y examina de qué manera el uso de la operación lógico-retórica de la consecuencia le permite al narrador construir una imagen del sí mismo como sujeto razonable.

La Parte IV, “Narrar en el molde de prácticas sociales”, incluye los capítulos 9, 10 y 11. En el capítulo 9, “Temporalidad narrativa específica de un género”, con un vasto corpus de narrativas producidas en alegatos finales en juicios penales orales, la autora demuestra que la combinación de ciertos recursos genera una ralentización de la acción narrativa y, con ello, se obtiene el efecto de una exposición de imágenes estáticas de los acontecimientos. El estilo resultante adquiere los valores de formalidad, profesionalismo y legitimación. El capítulo 10, “La experiencia ajena, el sentido común

y los relatos en cadena”, se trata de una única narración que ofrece la confrontación entre dos relatos fundados en tipificaciones culturales diferentes y la refutación del primero de ellos sobre la base del sentido común. Las expectativas asociadas en la *doxa* al estereotipo de una categoría de protagonista dotan de plausibilidad a una línea narrativa y se la restan a la contrapuesta. En el capítulo 11, “La incuestionabilidad de la experiencia personal”, la autora revela que, aún en un género institucional, se apela a narrativas mínimas de experiencia personal como estándar para evaluar los acontecimientos del relato principal. El concepto de “postura epistémica” posibilita explicar la construcción subjetiva de la información presentada.

La Parte V, “Atravesando acontecimientos institucionales de narración”, abarca los capítulos 12, 13 y 14. En los dos primeros, se trasciende el nivel del encuentro social para abordar las conexiones entre hechos de habla con sus ocasiones de narración. El último retoma los hilos de las discusiones precedentes y los integra para proponer considerarlos en los mundos digitales.

El capítulo 12, “Macro relato”, propone el modo de conceptualizar una narrativa completa –realizada por un receptor– que integra un conjunto de narrativas producidas por diferentes narradores, referidas a un mismo suceso y ordenadas alrededor de un “momento cero”. Isolda Carranza explora allí la asociación entre un cronotopo narrativo y el establecimiento de vinculaciones causales entre hechos del mundo diegético. En el capítulo 13, “Trayectorias de narrativas burocráticas”, la autora propone el concepto “situación de contacto” para caracterizar un encuentro social asimétrico (en la relación social y las diferencias de capital simbólico) como el que observa entre operadores judiciales y ciudadanos legos. Luego, centra su atención en el proceso de reelaboración y transformación del relato interaccional en un acta escrita y en la trayectoria institucional de esa historia, que será luego empleada selectivamente en otros acontecimientos comunicativos. El capítulo 14, “Interactuar y narrar en entornos cambiantes”, comienza con una mirada panorámica del recorrido del libro que subraya el ordenamiento de los nudos teóricos. Luego, proyectando una agenda de investigación, Carranza expone recientes desarrollos analíticos sobre narrativas digitales y reflexiona sobre los desafíos para tales estudios que presentan las redes sociales.

Sin duda, esta obra –caracterizada por el rigor metodológico y la claridad expositiva– es una valiosa contribución no solo para los estudios narrativos con perspectiva sociolingüística, sino para el campo del discurso interaccional y sus ámbitos sociales de proyección. Por otro lado, la relevancia social de los problemas que aborda, la profundidad del tratamiento y el compromiso ético de la investigadora hacen de esta publicación un modelo para los investigadores noveles y una referencia para los académicos del área.

Irania Malaver (Universidad de Granada). Reseña de Sáez, Daniel M.; Octavio de Toledo, Álvaro S. 2020. *Textos españoles de la primera mitad del siglo XVIII para la historia gramatical y discursiva. Vientos de arrastre y de cambio en la historia del español*. Madrid: Síntesis.

*Textos españoles de la primera mitad del siglo XVIII para la historia gramatical y discursiva. Vientos de arrastre y de cambio en la historia del español*¹ es una colección de documentos para el estudio del español de la primera mitad del siglo XVIII, organizada y editada por Daniel M. Sáez Rivera y Álvaro Sebastián Octavio de Toledo y Huerta.

La obra consta de un Prólogo, a cargo del catedrático emérito José Luis Girón Alconchel. Consta, asimismo, de la Introducción, en cuyo contenido extenso se exponen los criterios teóricos y metodológicos seguidos por los editores y, por último, de una sección tercera o sección final, en la que aparecen los sesenta documentos antologizados, organizados, a su vez, en dos partes:

- a) *Parte I Textos del primer cuarto del siglo XVIII (1701-1725)*, desde el texto 1 al 25, y cuya edición corre a cuenta de Daniel. M. Sáez Rivera.
- b) *Parte II. Textos del primer cuarto del siglo XVIII (1726-1758)*, editados por Álvaro S. Octavio de Toledo y Huerta y que van desde el texto 26 al 60.

Textos para la historia gramatical y discursiva anuncia un nuevo paso en un recorrido de investigación continuado y sólido que comenzó en 2002, cuando nació el [Proyecto de Investigación PROGRAMES: Procesos de gramaticalización en la Historia del Español](#)², adscrito a la Universidad Complutense de Madrid y cuyo interés se centra en el estudio del español del siglo XVIII con base en el análisis de textos no literarios, autores poco estudiados y en distintas *tradiciones discursivas* (TD) en el sentido en que son definidas y desarrolladas por Kabatek (2005).

El prólogo permite contextualizar este nuevo trabajo de Sáez Rivera y Octavio de Toledo y Huerta en el marco del proyecto de investigación fundado y dirigido por el catedrático Girón Alconchel y que se encuentra actualmente activo. El *PROGRAMES* tenía –y tiene– como fin “[...] completar la historia del español poniendo el foco, principalmente, en la transición del español clásico al moderno; para ello, la primera tarea no podía ser otra que recopilar los textos que nos sirvieran de base empírica para cumplir nuestro propósito” (pág. 13).

Sáez Rivera y Octavio de Toledo son miembros fundadores de ese *PROGRAMES* y, desde estos inicios, han continuado su labor de reunir textos que sirvan para el cometido de estudiar un período desatendido de la lengua. Los autores son investigadores especializados en la historia del español y en el estudio de procesos de cambio lingüístico. Además de reunir numerosas ponencias, artículos de investigación y participaciones en obras colectivas, Sáez Rivera y Octavio de Toledo y Huerta han editado otras antologías que se inscriben en la misma línea de impulsar la investigación gramatical-discursiva con corpus textuales que reflejen las distintas tradiciones discursivas y den cuenta de la –esperable– variación en los usos lingüísticos y los procesos de cambio de la lengua española. Es el caso de *En la estela del Quijote*:

cambio lingüístico, normas y tradiciones discursivas en el siglo XVII, del que Octavio de Toledo y Huerta es coeditor (Fernández Alcaide *et al* 2016) y de *Márgenes y centros en el español del siglo XVIII*, centrado en la descripción del español dieciochesco, coeditado por Marta Guzman Riverón y Sáez Rivera (Guzmán Riverón y Sáez Rivera, 2016).

La antología *Textos para la historia gramatical y discursiva* destaca, siguiendo a José Luis Girón Alconchel, porque se inscribe en la tradición investigadora hispánica y en la del hispanismo británico. En la primera, se encuentran don Ramón Menéndez Pidal con *Antología de prosistas españoles* y *Documentos lingüísticos de España* y *Lengua y literatura españolas medievales* de González Ollé. Y en el hispanismo británico, destacan los *Textos lingüísticos del medioevo español* de Gifford Hodcroft.

La claridad didáctica y la excelente presentación de los documentos son la segunda razón por la que destaca esta antología, que bien podrá ser de provecho tanto para los investigadores especializados del campo de la lengua y de la historia, como para el público culto general. Y, por último, para el catedrático, los documentos antologizados se suman a otros recursos para el estudio del español dieciochesco, especialmente, los contenidos en los recursos digitales. No menos importante son el compromiso y la ilusión con que los editores se han entregado a la búsqueda –en bibliotecas y archivos– de impresos y documentos del español del siglo XVIII: “un período de la historia del idioma –tan poco estudiado y tan poco conocido a principios del actual siglo XXI [...]”, de acuerdo con Girón Alconchel (pág. 13).

Sigue la Introducción, dedicada al plan de trabajo que sustenta la obra. Aquí los autores explican y argumentan exhaustivamente los criterios de selección y de edición de los documentos, aspectos esenciales en toda antología que se ofrece para estudios lingüísticos y, como se advierte, para estudios especializados más amplios. En las páginas dedicadas a la exposición de estos criterios encontrarán los lectores la ruta teórico-metodológica que se ha seguido: se explica, por una parte, la naturaleza textual de cada documento, la rica variedad textual que se ha alcanzado, tan necesaria para una nueva comprensión de la historia del español el siglo XVIII y, por la otra, la concepción sobre el estudio de lengua de la que parten los editores.

El contenido se subdivide en: **(1). Criterios de selección** y **(2). Criterios de edición y presentación de los textos**. En cuanto al punto **(1) Criterios de selección**, los editores exponen los tres ejes que sustentaron el proceso de selección y búsqueda en la etapa inicial de edición de la presente antología: las fuentes que se han tomado en cuenta para la búsqueda de los documentos, la etapa de la lengua que se aspira describir y la utilidad y relevancia de los documentos. Veamos, a continuación, los planteamientos de Sáez Rivera y Octavio de Toledo y Huerta sobre cada uno de estos criterios que, expuestos con un claro discurso científico a la vez que divulgativo, explican cómo se realizó la presente antología.

1.1 Beber en todas las fuentes; impresos y manuscritos en los albores de la modernidad

En la antología, se han reunido textos de fuentes impresas y manuscritas, fuentes transmitidas por imprenta mecánica o a mano, y textos diversos como epistolarios, relatos históricos o ensayos divulgativos. ¿Por qué textos impresos y manuscritos? Los impresos, al ser difundidos masivamente por la maquinaria editorial, permiten estudiar,

por una parte, las disposiciones de apoyo a la norma lingüística y, por la otra, las disposiciones conformadoras de dicha norma: “[...] los textos escritos con arreglo a la(s) norma(s) de prestigio, amplifican tal norma a través de la imprenta, la escuela, la Academia y otras entidades que pueden haber surtido material para la reflexión y la normalización académica” (pág. 16). Como muestra de este tipo de texto es el acta fundacional de la Real Academia de la Lengua (§ 29). En cuanto a los manuscritos, su relevancia se halla en que pueden contener o reflejar normas menos generalizadas, locales o regionales. Sáez Rivera y Octavio de Toledo y Huerta mencionan, por ejemplo, los escritos políticos, tales como testamentos de ministros o virreyes en los que se expresaba su ideario frustrado: “obras que buscan una rápida difusión mediante una proliferación de copias manuscritas tratando de esquivar, en la medida de lo posible, los escollos de la censura [...]” (pág. 16). Es el caso del texto §47, perteneciente a José del Campillo y Cosío.

Los editores van un paso adelante en la recolección de distintas TD: de una misma obra, se cuenta con una variante impresa (§ 26) y otra manuscrita (§46); también, de un mismo autor ofrecen una obra manuscrita (§14) y otra impresa (§16) y, por último, de un mismo autor, como por ejemplo del historiador Juan de Ferreras, ofrecen un fragmento impreso (§5) y una carta hológrafa (§55). Afirma Kabatek que “[...] cuando se estudia la historia de una lengua, lo que se estudia no es la evolución de la lengua [...] sino textos de diferentes épocas, textos que se consideran representativos de los respectivos estados de lengua” (2005:163). Así, *beber de todas las fuentes* es –en forma y fondo– una declaración de los editores sobre cómo concebir y emprender el estudio del cambio lingüístico a partir de los planteamientos de la lingüística coseriana que sustentan el enfoque de las TD (cf. Kabatek 2018).

1.2. Una antología para una época de vientos de arrastre y de cambio: el español del primer cuarto del siglo XVIII y la carencia de textos con que estudiarlo.

Sáez Rivera y Octavio de Toledo y Huerta intentan construir una muestra representativa para el estudio del español del siglo XVIII, un español que puede considerarse una frontera lingüística e histórica entre la época del español clásico del Siglo de Oro y la del español moderno. Y es así que los autores reflexionan: “[...] nos podemos preguntar si el español del siglo XVIII es clásico o ya moderno, o más bien algo entre medias [...], decisión que depende también de cómo definamos o consideremos lo clásico y lo moderno y su atribución a períodos de la lengua española” (pág. 17). Critican razonadamente la periodización de la lengua basada en la historia de la literatura, puesto que ha contribuido a ocultar la singularidad del español del siglo XVIII –clásico y moderno–, y llaman la atención sobre la necesidad de enriquecer los corpus textuales de este período. Y la ausencia de trabajos sobre autores o grupos de autores y sobre textos afines en tradición discursiva ha sido el resultado, prácticamente inevitable, de estos planteamientos.

Se menciona, en este apartado trabajos de mucho interés sobre morfosintaxis del español del 1660 al 1760. El lector encontrará un repertorio bibliográfico de mucha utilidad para emprender nuevas investigaciones a partir del estado de la cuestión que exponen Sáez Rivera y Octavio de Toledo y Huerta en estas páginas de la Introducción. Además de la valoración de estas investigaciones, hay una hoja de ruta de mucho valor

acerca de distintos aspectos del análisis lingüístico que necesitan desarrollarse, lo que hace crecer la pertinencia de esta antología.

1.3 Una antología para muchos públicos: utilidad múltiple y diversidad textual de este volumen.

En esta sección, los autores explican que han tenido en cuenta como destinatarios no solo a los historiadores de la lengua, sino también a historiadores generalistas y de la literatura. No muchos trabajos de este tipo, hechos con la rigurosidad que implica la búsqueda de selección de documentos para estudios lingüísticos, pueden ser útiles también para otros campos de investigación.

La antología contiene autores conocidos (reunidos en el segundo cuarto del siglo XVIII) y autores menos conocidos, desconocidos y marginales (primer cuarto del siglo XVIII). En cuanto a los documentos, se ofrece, también, como información sociolingüística relevante para estudios de dialectología y sociolingüística históricas, el lugar y fechas de nacimiento y muerte de los autores. Como bien señalan los editores desde el principio de la Introducción, la antología contiene textos de distintas *tradiciones discursivas* organizadas en el Cuadro 2 que muestra los universos del discurso, formas discursivas y temas discursivos de la antología (pág. 27)

Respecto al **punto (2)** acerca de los **criterios de edición y presentación**, los autores presentan una edición paleográfica que conserva las grafías, la acentuación y la puntuación originales. También se conservan las cursivas y el subrayado que también aparecen en los manuscritos. Para marcar sus intervenciones, Sáez Rivera y Octavio de Toledo y Huerta recurren a la etiqueta *metaindicaciones*, como por ejemplo la etiqueta *sic*, que emplean para indicar aquellos errores que pueden desvelar determinadas prácticas gráficas o usos fónicos. Esta explicación general se completa con una detallada descripción del sistema de abreviaturas que se emplea en la antología y que se presenta en el *Cuadro 3. Siglas y abreviaturas más frecuentes*, disponible para los lectores interesados.

Cada texto tiene un título abreviado, el nombre del autor en negrita y el lugar y año y muerte de nacimiento y muerte. Cuando se ha podido, una semblanza biobibliográfica del autor. Además de toda esta información, aparece el título completo en cursiva, lugar de impresión o copia, editor (o escribiente y año). Y, por último, los editores incluyen la localización del documento (biblioteca o archivo), una información muy relevante para todo investigador.

Para concluir, podemos afirmar que la colección de *Textos españoles de la primera mitad del siglo XVIII para la historia gramatical y discursiva*, de Daniel Sáez Rivera y Álvaro S. Octavio de Toledo y Huerta es un aliento para todos aquellos que se interesan por la historia de la lengua española, pues constituye un material de extraordinaria diversidad textual, editado con el rigor requerido y guiado por un enfoque gramatical-discursivo que aspira a comprender los hechos del pasado de la lengua en su compleja manifestación a través del “dato” multidiscursivo y multidimensional. Consecuentes con el planteamiento de que la lingüística histórica no puede obviar el estudio de la variación de las tradiciones discursivas, además del estudio de las variaciones sociodialectales y diafásicas (Kabatek 2008), *Textos españoles de la primera mitad del siglo XVIII para la historia gramatical y discursiva. Vientos de arrastre y de cambio en*

la historia del español auguran nuevas y necesarias investigaciones sobre la lengua española del siglo XVIII.

Referencias bibliográficas

Fernández Alcaide, Marta; Leal Abad, Elena; Octavio de Toledo y Huerta, Álvaro S., eds. 2016. *En la estela del Quijote. Cambio lingüístico, normas y tradiciones discursivas en el siglo XVII*. Bern: Peter Lang.

Guzmán Riverón, Martha; Saéz Rivera, Daniel M., eds. 2016. *Márgenes y centros en el español del siglo XVIII*. Valencia: Tirant Humanidades.

Kabatek, Johannes, ed. 2008. *Sintaxis histórica del español y cambio lingüístico. Nuevas perspectivas desde las tradiciones discursivas*. Frankfurt/Madrid: Iberoamericana/Vervuert.

Kabatek, Johannes. 2005. Tradiciones discursivas y cambio lingüístico. *Lexis* 29.2: 151-177.

Notas

¹ También “*Textos para la historia gramatical y discursiva*”.

² <https://www.ucm.es/procesosdegramaticalizacionenlahistoriadelespanol/documentos-programas>

Alba Nalleli García Agüero (Universität Bern). Reseña de Mancera, Ana; Pano, Ana. 2020. *La opinión pública en la red. Análisis pragmático de la voz de los ciudadanos*. Madrid / Frankfurt: Iberoamericana / Vervuert.

Indudablemente, la última década ha sido marcada por la aparición y desarrollo de las tecnologías 2.0. Estas, conocidas como web social o web 2.0, han dado lugar a un cambio radical en la manera en que el usuario de internet se relaciona con la información proporcionada: el destinatario ha dejado de ser un simple receptor, consumidor de información para convertirse en productor y aportador de la misma. La web 2.0 se ha convertido, pues, en una plataforma de trabajo colaborativo e intercambio de datos y opiniones. Evidentemente, estos entornos digitales han propiciado un espacio mucho más abierto al debate colectivo y a la opinión pública en contraste con los medios informativos tradicionales. Dichas posibilidades han tenido un fuerte impacto social, que se puede considerar tanto positivo como negativo. Así, por ejemplo, se han hecho posibles la organización y movilización política, el intercambio de puntos de vista, pero también se han fomentado la manipulación de información, el ataque directo e inclemente a ideas y personajes, así como la polarización de opiniones.

La opinión pública en la red de Ana Mancera Rueda y Ana Pano Alamán adopta las perspectivas del análisis del discurso y de la pragmática lingüística para analizar las distintas manifestaciones de la opinión pública en la web social. A través de distintos textos recabados de Twitter, Facebook, Change.org, *blogs* de autor, Amazon, YouTube, foros de discusión, TripAdvisor e Instagram, las autoras develan las estrategias discursivas y los recursos lingüísticos más utilizados por los usuarios de internet para expresar su opinión y, muchas veces, para “remover conciencias” e influenciar la manera de pensar y comportarse de otros cibernautas. La obra está dividida en 12 capítulos, escritos de manera alternada por las autoras. Cada capítulo trata uno de los medios digitales, abordando el tema a partir de un aspecto teórico discursivo o pragmático.

Así, en el capítulo 1, Pano Alamán reseña la concepción y los matices que ha ido adquiriendo la idea de “opinión pública” desde la Antigüedad hasta la llegada de Internet. También destaca las principales características de los medios sociales concentrándose en la “conversación política” que se ha generado a través de ellos. Aunque se han depositado muchas esperanzas en el internet como un medio capaz de fomentar la democracia mediante la conversación política informal, Pano Alamán devela algunos factores que impiden que se lleven a cabo acciones políticas realmente efectivas. Debido a que la esfera pública virtual es un espacio abierto a los sentimientos y la subjetividad colectivos, el capítulo habla también de los medios que se utilizan para recoger y medir sentimientos, opiniones y actitudes en el ciberespacio. Aunque los algoritmos utilizados son bastante efectivos, la autora advierte que su eficacia disminuye cuanto más se centran en fenómenos lingüísticos y pragmáticos complejos como la ironía y el miedo. Para finalizar el capítulo habla de la “otra cara de la moneda”, es decir, de la posibilidad que tienen los medios digitales para controlar datos y manipular opiniones. Finalmente, la autora hace referencia a los temas que se han tratado dentro del análisis lingüístico y pragmático de la expresión de la opinión pública en la red.

En el capítulo 2, Pano Alamán habla de cómo las tecnologías digitales y la red 2.0 han multiplicado las posibilidades de los ciudadanos para organizarse con otros en torno a una misma causa y proponer iniciativas a sus gobiernos para cambiar un estado de cosas. Hace alusión de dos tipos de plataformas en línea que se han creado para promover iniciativas ciudadanas en contextos locales, nacionales e internacionales: aquellas que buscan crear grupos de interés y movilizarse en favor de una causa (por ej. MoveOn) y aquellas plataformas que sirven para tramitar quejas y solicitar a las instituciones la resolución de algún problema local. La autora dedica el capítulo a tres plataformas de este último tipo: Tuderechoasaber, Irekita e ICE. Comienza explicando el funcionamiento de Tuderechoasaber, primera web española que permitió pedir información a cualquier institución pública de España. De esta plataforma, analiza cómo los usuarios construyen sus peticiones. A pesar de que este recurso ha sido muy utilizado, el sistema es muy similar al buzón de quejas tradicional, en el que no se pueden consultar los mensajes de los usuarios ni el seguimiento del trámite. No es el caso de la web Irekita, que sí permite la interacción directa entre los ciudadanos y el gobierno vasco. Pano Alamán explica cómo funciona esta página y la estructura de las peticiones que, además, son públicas y están abiertas al debate y al voto. Asimismo, habla del registro oficial en línea de Iniciativa Ciudadana Europea (ICE) de la Comisión Europea, la cual contiene una lista de iniciativas abiertas para la recogida de declaraciones de apoyo. Además de describir lo que contiene la página web, muestra ejemplos de las iniciativas que, a diferencia de las aparecidas en Tuderechoasaber, formuladas de manera formal, las de la ICE reflejan más inmediatez comunicativa, llegando incluso a resultar descorteses. La autora muestra con un ejemplo las partes que componen los formularios que permiten enviar las iniciativas, así como la organización de los textos de la petición. Desde la perspectiva de la gramática textual (*cfr.* Calsamiglia y Tuson 1999; Beaugrande y Dressler 1997; Marimón Llorca 2008), Pano Alamán expone las relaciones de coherencia y cohesión que aparecen en los textos. Por ejemplo, presenta ejemplos con los que muestra cómo la reiteración léxica; el uso de *ello/por ello* como mecanismo de referencia anafórico; la elipsis (verbal y nominal); la estructura informativa S-V-OD-OI; la topicalización; la enumeración; y el uso de marcadores discursivos son mecanismos de cohesión muy habituales en este tipo de textos.

El capítulo 3, escrito por Mancera Rueda, se centra en la plataforma de movilización ciudadana Change.org. Narra sus orígenes y la unión con la plataforma española para campañas ciudadanas, llamada Actuable. Asimismo, describe los pasos y sugerencias propuestas por la misma plataforma para redactar una petición con suficiente impacto, que funja como instrumento de presión y movilice a las personas e instituciones pertinentes. La parte central del capítulo da cuenta de los recursos argumentativos más utilizados por los usuarios de la plataforma para alcanzar sus objetivos. En efecto, a partir del análisis de ochenta peticiones publicadas en Change.org, la autora devela las constantes que aparecen en la organización macroestructural de los textos. Analiza las estrategias gramaticales empleadas en los títulos, por ejemplo, el uso del imperativo con sujeto tácito, la introducción del título mediante el adverbio de negación *No*, la apelación a los lectores a través de la segunda persona plural (“No os olvidéis de...”) o el *nosotros* inclusivo para involucrar y dar responsabilidad al lector, etc. También habla de las estrategias que otorgan fuerza argumentativa para convencer al lector de apoyar la iniciativa: la secuencia en que están organizados los textos, la utilización de tecnicismos, símiles, metáforas, testimonios, la apelación a la autoridad, la sorpresa,

topoi, la ironía, enunciados parentéticos, adjetivos calificativos intensificadores, etc. Además, la autora devela que existen constantes en el cierre de la petición, donde se le pide al internauta su firma. Finalmente, se exponen algunas polémicas y críticas sobre la credibilidad y el sistema de financiación de la plataforma.

Pano Alamán, dedica el capítulo 4 al foro de discusión, espacio digital que desde los inicios de internet fue pensado para albergar las opiniones de los usuarios. Como en el capítulo anterior, en este se comienza por los orígenes de este medio, se mencionan sus precursores, se describen sus características y se hace referencia a los diferentes tipos de foros. Además de los foros, Pano Alamán también alude a otro tipo de espacios virtuales que han ganado fuerza en los últimos años, y que se han denominado como “comentarios a las noticias”. La autora compara estos textos con las tradicionales “cartas al director” por su forma y función: son textos argumentativos, relativamente cortos, en los que los lectores reaccionan ante una noticia y la comentan, pero a diferencia de las cartas, la posibilidad de reacción es inmediata y potencialmente dialógica. Tras haber descrito estas dos “ágoras virtuales”, los foros y los comentarios en la prensa digital, la autora hace una comparación entre ellos y enuncia sus diferencias (número de caracteres, manera en que se archivan los mensajes, etc.) y similitudes (la naturaleza polilógica de las interacciones, la brevedad, estilo coloquial). Asimismo, destaca que las características de estos espacios imponen la utilización de ciertas estrategias discursivas que garanticen la coherencia y el carácter persuasivo y argumentativo de los textos. En específico, refiere el uso del vocativo y del discurso referido. A partir de la teoría de Ducrot (1984) sobre la multiplicidad del sujeto enunciativo y de los conceptos de intertextualidad y heteroglosia (Bajtin 1981), la autora analiza la manera en que las diferentes voces se hacen oír y se van entrelazando en el discurso desarrollado en estos espacios digitales. Así, tomando como objeto de análisis 160 mensajes de cuatro foros de debate y 225 comentarios sobre nueve noticias publicadas en las ediciones digitales de *El País*, *El Mundo* y *ABC*, ilustra que las estrategias más frecuentes para integrar los discursos son el estilo directo, el indirecto y el indirecto libre, y explica con los ejemplos cuáles son los objetivos que se pretenden alcanzar con cada uno de estos procedimientos de citación. Por último, hace referencia a otras estrategias también recurrentes en este tipo de espacios con las que los internautas intentan transmitir sus opiniones, a saber, la incorporación de extractos literarios, así como de frases, modismos y refranes que apelan al conocimiento compartido y la inserción de otros códigos y variedades lingüísticas (heteroglosia).

En el capítulo 5, Mancera Rueda aborda el tema de los “grandes almacenes virtuales” E-Bay y Amazon, los cuales también propician un espacio para la expresión de la opinión pública. De hecho, según la autora, las opiniones y reseñas adquieren gran importancia para las plataformas, pues más que los precios o las fotos de los productos, el dictamen de otros consumidores es determinante para la adquisición de los productos en venta. Por esta razón, Amazon, por ejemplo, incentiva con obsequios a los críticos que adquieren prestigio; algunos comerciantes contratan internautas que redacten críticas híper-positivas que coloquen sus productos entre los que tienen cinco estrellas; o incluso han llegado a ofrecer reembolso a los clientes a cambio de reseñas positivas, aunque esta acción sea prohibida por las plataformas. Las opiniones falsas han motivado la creación de modelos matemáticos que pueden identificar los productos que se benefician con tales valoraciones. A través de la teoría de los actos de habla (Searle 1986), Mancera Rueda analiza un corpus de 200 valoraciones de productos en estas

plataformas y da cuenta de los tipos de actos que predominan. Así pues, dilucida que los actos que más aparecen son los asertivos, los cuales brindan un valor veritativo a la argumentación. Tales actos son expresados mediante diferentes estrategias pragmáticas como locuciones performativas atenuadas para expresar la opinión del evaluador; predicados doxásticos introducidos por el verbo “creer”, predicados dubitativos que mitigan la afirmación, el “se” impersonal, el nosotros inclusivo y el tú impersonal que desfocalizan al hablante, etc. Otro tipo de actos recurrente es el de los actos directivos en su modalidad de consejo expresado de distintas maneras (recomendaciones basadas sobre la experiencia, especulaciones, guías sobre cómo actuar, etc.). Mancera Rueda también reconoce la frecuencia de los actos comisivos con los que el evaluador promete no volver a comprar el producto. Y finalmente, actos expresivos cuya fuerza ilocutiva es la de expresar emociones como arrepentirse, disculparse y agradecer.

En el capítulo 6, Pano Alamán analiza los blogs llamados “de autor” publicados en ediciones digitales de distintos medios. Introduce el tema exponiendo la historia de estos recursos, la descripción de sus características y la mención de los tipos existentes. Como afirma la autora, estas “bitácoras digitales” son un excelente medio para el aporte de contenidos actualizados y para el intercambio de opiniones y comentarios entre los blogueros y sus lectores. En efecto, en este medio la audiencia cobra gran protagonismo ya que no solo comenta, sino que sus exigencias influyen en los contenidos del blog. En este sentido, Pano Almán dice que, para ganar un mayor número de seguidores, los blogueros recurren a estrategias lingüísticas con el fin de captar la atención y mantener el interés de sus lectores. Entre estas estrategias se encuentra el uso del registro informal, el cual se logra a través de enunciados interrogativos y apelativos, así como de mecanismos de atenuación e intensificación. Es justamente en los mecanismos intensificadores que la autora fija su atención y los analiza en 140 entradas de la versión digital de cinco diarios españoles en los que los autores escriben sobre cine, series, arte o literatura. Así, basándose en la noción pragmalingüística de *intensificación* (Briz Gómez 1998, Albelda Marco 2005), reconoce los recursos morfemáticos, léxicos, sintácticos y semánticos con los que se manifiesta esta categoría. Por ejemplo, encuentra el uso frecuente de sufijos como *-ísimo/a*, *-azo*, *-ote* y en menor grado el prefijo *súper-*. También da ejemplos que contienen sustantivos, adjetivos, verbos y adverbios cuyos semas implican un valor intensificador (*quintaesencia*, *apabullante*, *sobrecogedor*, *fulminantemente*, etc.), así como distintos recursos fraseológicos, tales como locuciones adnominales, locuciones verbales y adverbiales que contienen cierto grado de idiomatización. Otros recursos intensificadores de naturaleza semántica observados en el corpus son la metáfora, la hipérbole, expresiones más o menos fijadas, referencias a la cultura y a las catástrofes, los cuales contribuyen de manera creativa a adjudicar un valor agigantado al producto o a los sentimientos que este provoca. Desde el punto de vista de la modalidad, la autora encuentra que aparecen verbos performativos como *jurar* y *decir*; enunciados exclamativos y exhortativos; así como partículas discursivas con valor de evidencialidad, todos ellos recursos que intensifican la actitud del autor y expresan un alto grado de modalidad epistémica. Finalmente, la autora explica y ejemplifica las funciones de la actividad intensificadora (funciones autorreafirmativa, alorreafirmativa y contrarreafirmativa (Briz Gómez 2016).

El capítulo 7 está dedicado al análisis de reseñas sobre hoteles y restaurantes publicados en portales de opiniones en línea. Dichos portales como TripAdvisor, Booking o Trivago han revolucionado la comunicación en el sector turístico puesto que se ha

llegado a afirmar que las opiniones que estos albergan sostienen a los sectores hotelero y gastronómico. En efecto, las reseñas hechas mediante estos portales han adquirido gran importancia comercial, ya que las valoraciones influyen significativamente en las decisiones de los potenciales clientes, así como también pueden arruinar la reputación de las empresas turísticas. Conscientes de los efectos negativos, afirma Mancera Rueda, muchos internautas utilizan estrategias de atenuación lingüística al momento de hacer una evaluación. La autora, pues, se enfoca en el reconocimiento de las estrategias mitigadoras más utilizadas en estos entornos digitales. Parte exponiendo teóricamente el fenómeno pragmalingüístico de la *atenuación* (Briz Gómez 1995, 1998) para después pasar a la ejemplificación de los recursos mitigadores en diferentes reseñas. De este modo, devela que los procedimientos atenuantes más utilizados son el uso de diminutivos, la cuantificación minimizadora (“un poco”), difusores significativos (“algo”), cuantificadores aproximativos (“más o menos”) y extranjerismos (*top*). Los recursos anteriores minimizan o “desdibujan” el contenido proposicional de lo asertado, pero también hay estrategias discursivas y elementos lingüísticos con los que se reduce la fuerza ilocutiva de los actos de habla, por ejemplo, mediante el uso de verbos modales en condicional o con actos asertivos de opinión (“a mi gusto”), adverbios y locuciones adverbiales con los que se muestra duda e incertidumbre, reduciendo el compromiso epistémico con la verdad (“quizá”, “tal vez”). Dicha responsabilidad de la producción del acto de habla se suele también reducir mediante el uso de formas impersonales, estructuras adversativas y concesivas, así como de locuciones adverbiales exceptivas que buscan restringir el alcance negativo de la crítica al mismo tiempo que cuidar las imágenes del autor y del negocio. Interesante es el hallazgo de la frecuencia con que el opinante se siente obligado a insertar una atenuada crítica negativa para resultar más creíble. Mancera Rueda, además de analizar reseñas de clientes, también aborda de la misma manera los comentarios que los propietarios de los negocios dejan como respuesta a sus evaluaciones. Expone que se recurre a actos de habla expresivos (agradecimientos) y comisivos, así como recursos atenuadores que muchas veces contribuyen a subrayar aspectos positivos de la empresa en beneficio de su imagen positiva.

En el capítulo 8, Pano Alamán trata la red social más usada en los países de habla hispana y quizá del mundo entero: YouTube. Esta plataforma digital, cuyo objetivo es publicar y compartir vídeos, contiene millones de datos no solo audiovisuales, sino también textuales y es de suma utilidad para las empresas y la promoción de sus productos. La autora introduce el tema mencionando los vídeos más vistos entre 2018 y 2019 en España y los canales con más suscriptores. Prosigue con los orígenes y el desarrollo histórico de este entorno digital y hace una descripción de sus características a nivel comunicativo. También resuelve la pregunta sobre por qué algunos vídeos son tan visitados mientras otros no causan interés. Esto se debe principalmente al algoritmo que constituye YouTube, el cual determina los resultados de las búsquedas de los usuarios, así como las notificaciones y publicidades que reciben. Además, habla sobre las polémicas que se han suscitado sobre esta plataforma a partir de testimonios de desarrolladores que han trabajado en ella, por ejemplo, que está pensada para promover la violencia y el conflicto. Debido a que los contenidos de este tipo de plataformas determinan el comportamiento y orientación de los usuarios y fomentan la polarización de la opinión pública, la autora considera muy oportuno hacer un análisis del discurso de los comentarios publicados en este medio. Antes de comenzar con su análisis presenta una lista muy atinada de autores que han estudiado

los contenidos de esta plataforma desde la lingüística y la pragmática en el ámbito hispánico y anglosajón. Asimismo, expone el modelo de Koch y Österreicher (1985), el cual propone que las modalidades de uso lingüístico se encuentran en un *continuum*, cuyos polos son la inmediatez y la distancia comunicativa. Sobre esta base teórica y apoyándose también en la clasificación del grupo Val.Es.Co, la autora afirma que los enunciados expresados en esta red social suelen situarse en el polo de la inmediatez o en el campo de la *realización coloquial* (Briz Gómez 2010). Tal afirmación es sustentada mediante el análisis de 600 comentarios publicados en relación con seis vídeos. Los rasgos coloquiales que la autora distingue a nivel fónico son, por ejemplo, las faltas de ortografía en “heterografías” (desviaciones intencionales de la norma), la falta de puntuación y tildes, las mayúsculas para intensificar, la reiteración de vocales, emoticonos, etc. En el plano morfosintáctico, reconoce la creación de palabra por sufijación, la arroba para integrar masculino y femenino, diminutivos, fusión de palabras, locuciones adjetivas y adverbiales introducidas por *de*, interjecciones, etc. A nivel sintáctico devela la recurrencia de la yuxtaposición y la coordinación en detrimento de la subordinación, el uso de puntos suspensivos y la elipsis. En el plano léxico detecta expresiones disfémicas, neologismos despectivos, palabras del argot juvenil y préstamos del inglés, entre otros.

El capítulo 9, escrito por Mancera Rueda, se centra en la cuarta red social más utilizada en España (después de WhatsApp, Facebook y YouTube), según estudios oficiales: Instagram. La autora analiza los mensajes que reciben algunos personajes de cierto renombre en España a partir de las fotografías y vídeos que comparten en dicha plataforma. Específicamente, se centra en las felicitaciones que reciben por parte de sus seguidores, por lo que se apoya en la noción pragmática de *cortesía verbal* para develar los mecanismos lingüísticos más utilizados para comunicar sus halagos. Comienza, por tanto, mencionando a los autores más destacados que han desarrollado teorías sobre la cortesía verbal (Lakoff 1973; Fraser 1980; Leech 1983; Brown y Levinson 1978; Lavandera 1988; y Kerbrat-Orecchioni 2000) y la característica distintiva de sus propuestas. Profundiza un poco más en la noción de *imagen social* desarrollada por Goffman (1967) y retomada en el ámbito hispánico por Bravo (2005) y Fuentes Rodríguez (2010a). Debido a que este tipo de redes sociales promueven claramente el trabajo de autoimagen, muchos analistas del discurso han puesto su atención en ellas y en la construcción identitaria que suscitan. La autora presenta algunos de estos estudios antes de pasar a su análisis. Instagram es la red social que más seguidores ha ganado en los últimos dos años y ha sido muy utilizada por las celebridades españolas para comunicar su trabajo, sus logros y reconocimientos, por lo que Mancera Rueda comienza su análisis a partir de las publicaciones de estas figuras públicas. Expone que los mensajes conllevan claramente actividades en las que las celebridades moldean positivamente su imagen mostrándose como personas de éxito, pero también como personas solidarias que reconocen los logros de sus compañeros. A este punto, la autora hace una útil reseña sobre las nociones de imagen positiva y negativa de Brown y Levinson (1978) y las contrasta con las nociones más actualizadas de *autonomía* y *afiliación* trabajadas por Scolon y Scolon (1995) y, en el ámbito del español por Bravo (1999) y Boretti (2002). En esta línea, la autora demuestra que en Instagram las actividades de imagen que predominan son de tipo afiliativo. Al mostrar aprecio por sus colegas y seguidores, los famosos reciben de parte de estos, actos propios de la *cortesía valorizante* (Kerbrat-Orecchioni 1996), mediante los cuales se refuerza su imagen. La mayor parte de las respuestas de los fans son muestras de *cortesía ritual* (saludos,

felicitaciones, etc.), en los que destacan marcadores discursivos que acortan la distancia interpersonal y elementos que refuerzan la imagen positiva de la celebridad. Los cumplidos son predominantes en las publicaciones, y se expresan a través de enumeraciones adjetivales, superlativos y elementos intensificadores (alargamientos vocálicos, sufijos, adverbios cuantificadores, escritura *ideofonemática*, etc.). También se encontraron manifestaciones de *cortesía estratégica* (Fuentes Rodríguez 2010b), es decir, aquella que se usa para alcanzar un objetivo argumentativo claro como solicitar algo de las celebridades. Evidentemente, estos actos resultan amenazadores, por lo que los fans recurren a distintas estrategias de atenuación.

El capítulo 10 está dedicado a la red social número uno, colocada a la cabeza por la alta cantidad de usuarios (2.271 millones): Facebook. La autora del capítulo, Pano Alamán, toma como objeto de análisis los cumplidos que se generan en esta plataforma. Comienza describiéndola. Menciona, por ejemplo, que tal medio se ha vuelto en una red principalmente visual que impone una organización de la línea de tiempo basada en el ritmo de la publicación de fotos, vídeos o memes. El fin de tales imágenes es provocar reacciones en forma de comentarios, clics del dispositivo *Me gusta*, o *tags* que etiquetan personas y lugares. También afirma que, así como Instagram, Facebook se ha considerado una red adulatoria en la que abundan comentarios halagadores, lo cual influye en el ánimo y autoestima de los usuarios. En vista de que generalmente la intención comunicativa es propiciar la reacción de los interlocutores mediante preguntas y petición de consejos y sugerencias, la autora sostiene que son muy frecuentes los enunciados exhortativos, interrogativos, el modo imperativo, la segunda persona del singular y del plural y los vocativos, así como el uso de elementos pertenecientes al registro oral coloquial. Asimismo, Pano Alamán habla del impacto emotivo del *me gusta*, el cual es un indicador del interés que la publicación puede provocar. Remite también a los “tiempos” que tienen lugar en la plataforma en virtud de los distintos tipos de interacción que esta posibilita, y del algoritmo Edgerank, el cual determina el orden en que aparecen las publicaciones según la afinidad y grado de interacción entre los “interlocutores”. Otro de los elementos de la interfaz de Facebook mencionados por la autora son los *emojis* que, junto con el *me gusta*, han sido un instrumento para medir las opiniones de los usuarios y llevar a cabo estudios de mercado. Antes de iniciar con su análisis, hace una descripción del cumplido como acto de habla, de las funciones que desempeña, menciona los actos de habla que suelen acompañarlo y la clasificación que se ha hecho de estos en la lingüística hispánica, así como los estudios que se han realizado de ellos en entornos digitales. Asimismo, hace referencia a las formas lingüísticas mediante las cuales se suelen expresar y que han sido identificadas por otros estudiosos. Así, mediante el análisis de diferentes perfiles en Facebook (no menciona cuántos, ni cuáles fueron los parámetros para su selección), la autora detecta que las estructuras de los cumplidos coinciden con las encontradas en otros estudios. Por ejemplo, el uso de exclamaciones iniciadas con “que”; cláusulas declarativas que implican el proceso mental de “gustar” o “encantar”; la elisión del verbo generalmente copulativo; interjecciones, locuciones interjectivas y expresiones semifijas; léxico coloquial; ortografía prosódica intensificadora; *emojis* y *gifs*, etc. Finalmente, Pano Alamán demuestra cómo los aspectos que se valoran positivamente dependen del tipo de destinatario y del contexto situacional en el que se desenvuelve o al que pertenece (político, artístico, empresarial, etc.).

Mancera Rueda se encarga del capítulo 11 donde trata el tema de la descortesía verbal hacia periodistas en la red social Twitter. Comienza exponiendo ejemplos de tuits generados en España y en otras partes para mostrar el grado de agresividad que estos comentarios pueden vehicular. Asimismo, ofrece estadísticas de la cantidad de tuits ofensivos que las mujeres periodistas reciben en Estados Unidos y Reino Unido y de la propensión que tienen las mujeres de raza negra y de las que se inclinan por una ideología de derechas a ser las principales víctimas de estos ataques. En seguida, la autora prosigue con la parte teórica donde menciona autores que han tratado el tema de la descortesía verbal, entendida no como un fenómeno subordinado a la cortesía, sino como un fenómeno en sí que comprende estrategias encaminadas a minar la imagen del interlocutor (Lakoff 1989; Culpeper 1996; Martín Rojo 2000; Haverkate 2001; Placencia 2001). Más adelante expone cómo el trabajo de imagen que llevan a cabo los periodistas corresponde con su rol social y cómo esto es blanco de ataques. Tales agresiones, de acuerdo con el análisis de la autora, se expresan, por ejemplo, mediante preguntas retóricas; bombardeo de preguntas; críticas a su desempeño profesional; insultos; cuestionamientos a sus roles sociales, y otros actos amenazadores de la imagen, tanto del periodista como de la imagen de sus familiares.

En el capítulo 12, último capítulo del libro, Mancera Rueda analiza el fenómeno viral de los memes y la función y consecuencias que tienen en las discusiones por internet. La autora comienza con la explicación del origen del término *meme* acuñado por Richard Dawkins en 1976 (idea o comportamiento que se expande de persona a persona dentro de una determinada cultura, así como suelen hacerlo los genes y los virus). Ejemplifica cómo algunos memes han sido falsos en la sociedad norteamericana y los comportamientos que han conllevado. Asimismo, expone las ideas de Rodríguez Marín (2013) de que la supervivencia de un meme depende de su correlación con otros y de que vivimos en una *memeocracia* favorecida por el surgimiento de internet donde los cibernautas son transmisores de estos virus mentales. Antes de pasar a la parte teórica, Mancera Rueda habla de cómo, en los años noventa, internet se apropió del término de Dawkins para referirse a una serie de imágenes y vídeos acompañados por un breve texto que suele cambiar a voluntad del usuario. Prosigue con una breve exposición de aspectos relevantes de la teoría del humor verbal de Attardo y Raskin (1991) así como del modelo pragmático propuesto por el Grupo de Investigación sobre la Ironía y Humor en Español de la Universidad de Alicante. Después de esto, presenta su análisis sobre memes de contenido político difundidos recientemente mediante Twitter. Así, ejemplifica cómo, según la teoría de Raskin (1985), el humor se logra superponiendo dos guiones antitéticos y, naturalmente, conociendo la situación comunicativa. En este sentido, advierte también que para que un meme sea viral, es necesario que el contenido que evoca sea fácilmente reconocible por los internautas, de ahí que abunden los ejemplos de memes en los que se contrasta la escena política con el ámbito cinematográfica o televisivo. Mancera Rueda devela también que la parodia es otro recurso humorístico frecuente y da ejemplos de ello. Da cuenta de que los creadores de los memes recurren a varios tipos de tipologías narrativas, por ejemplo, en algunos destacan las secuencias descriptivas, mientras que en otros, el texto y la imagen están al servicio de la narración. Finalmente, la autora recuerda que para que el lector pueda interpretar el humor y esto tenga un efecto perlocutivo (la propagación del meme) el destinatario debe contar con una triple competencia: la lingüística, la genérica y la ideológica (Hutcheon 1994).

Queremos terminar esta reseña recordando que la web social está presente en el día a día de millones de personas. Muchas de estas participan de manera activa en la producción de información en la red, otras son más pasivas y solo reciben, pero todas están en contacto a través de la lengua y otros sistemas semióticos y pueden llegar a modificar sus ideas y comportamientos a partir de las opiniones navegantes. Porque como lo advierten las autoras, la sencillez, gratuidad y rapidez para comunicar mensajes que la web social ha favorecido, ha dado voz a colectivos ajenos a las élites. Además, la voz pública en la red ha modificado todo el sistema de comunicación que tienen las empresas, las instituciones, y hasta los actores políticos con sus destinatarios. Por estas razones, conocer el funcionamiento y efectos de la opinión pública en los medios digitales resulta de gran importancia e interés. El principal valor de este monográfico radica precisamente en ser uno de los pocos trabajos en el ámbito hispánico que se han abocado en dar luz a este fenómeno. Además de lo anterior, podemos decir que el libro de Mancera Rueda y Pano Alamán tiene otros puntos fuertes. En primer lugar, ofrece lo prometido en la introducción: narra los orígenes de cada uno de estos espacios digitales, esclarece la manera en que funcionan, la medida en la que influyen en la sociedad, las estrategias retóricas que los cibernautas utilizan para alcanzar sus objetivos comunicativos y buscan influir en el comportamiento de otros usuarios. En segundo lugar, presenta la explicación de algunos términos y teorías propios de la pragmática lingüística a partir de los cuales las autoras hacen el análisis de los ejemplos. Aunque se trata más bien de una exposición somera de los temas teóricos, esta ofrece referencias bibliográficas útiles y adecuadas para quien se interese por profundizar en estos temas. En este sentido, como lo manifiestan las autoras, el monográfico puede ser muy útil para analistas del discurso, expertos en *marketing* digital, periodistas enfocados en información digital y para estudiantes que aspiren a especializarse en estos ámbitos. Si tenemos que hablar de aspectos de este manuscrito que se podrían reforzar, podemos decir que habría sido interesante que las autoras presentaran un análisis cuantitativo y/o comparativo con medios más tradicionales para poder ver claramente en qué medida “abundan” (como sostienen en todos los capítulos) los recursos discursivos encontrados en la web social o que son más propios de esta.

Referencias bibliográficas

- Albelda Marco, M. 2005. *La intensificación en el español coloquial*. Tesis doctoral. Valencia: Universidad de Valencia.
- Attardo, S. y Raskin, V. 1991. Script theory revis(it)ed: Joke similarity and Joke representation model. *Humor*, 4 (3-4): 293-347.
- Bajtin, M. 1981. *The Dialogic Imagination*, Austin: University of Texas Press.
- Beaugrande, R. de y Dressler, W. 1997. *Introducción a la lingüística de texto*. Barcelona: Ariel.
- Boretti, S. H. 2002. Cortesía, imagen social y contextos socioculturales en la variedad del español de Rosario, Argentina. En D. Bravo ed. *Actas del Primer Coloquio Internacional del Programa EDICE “La perspectiva no etnocentrista de la cortesía: identidad sociocultural de las comunidades hispanohablantes”*. Stokolm: Programa EDICE, Stokolm Universitet, pp. 98-108.

- Bravo, D. 1999. ¿Imagen “positiva” vs. Imagen “negativa”? Pragmática socio-cultural y componentes de face. *Oralia: Análisis del discurso oral*, 4: 299-314.
- Bravo, D. ed. 2005. *Estudios de la descortesía en español. Categorías conceptuales y aplicaciones a corpora orales y escritos*. Buenos Aires: Editorial Dunken.
- Briz Gómez, A. 1995. La atenuación en la conversación coloquial. Una categoría pragmática. En L. Cortés Rodríguez ed. *El español coloquial: Actas del I Simposio sobre análisis del discurso oral*. Almería: Universidad de Almería, pp. 103-122.
- Briz Gómez, A. 1998. *El español coloquial. Esbozo de pragmatogramática*. Barcelona: Ariel.
- Briz Gómez, A. 2010. El registro como centro de la variedad situacional. Esbozo de la propuesta del grupo Val.Es.Co. sobre las variedades diafásicas. En I. Fonte y L. Rodríguez Alfano coords. *Perspectivas dialógicas en estudios del lenguaje*. Monterrey: Universidad Autónoma de Nuevo León, pp. 21-56.
- Briz Gómez, A. 2016. Evidencialidad, significados pragmáticos y partículas discursivas en español. Sobre la intensificación como significado evidencial. En R. González Ruiz; D. Izquierdo y Ó. Loureda Lamas eds. *La evidencialidad en español: teoría y descripción*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert.
- Brown, P. y Levinson, S. 1978. *Politeness. Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Calsamiglia, H. y Tuson, A. 1999. *Las cosas del decir. Manual de análisis del discurso*. Barcelona: Ariel.
- Culpeper, J. 1996. Towards an anatomy of impoliteness. *Journal of Pragmatics*, 25: 349-367.
- Ducrot, O. 1984. *El decir y lo dicho. Polifonía de la enunciación*. Barcelona: Paidós.
- Fraser, B. 1980. Conversational mitigation. *Journal of pragmatics*, 4: 341-350.
- Fuentes Rodríguez, C. 2010a. Ideología e imagen: ocultación en la prensa de la violencia social o lo políticamente correcto. *Discurso y Sociedad*, 4 (4): 853-892.
- Fuentes Rodríguez 2010b. *La gramática de la cortesía en español/LE*. Madrid: Arco Libros
- Goffman, E. 1967. On face-work. An analysis of ritual elements in social interaction. En E. Goffman ed. *Interactional Ritual. Essays in face to face behavior*. New York: Anchor Books, pp. 5-45.
- Haverkate, H. 2001. Cortesía y descortesía en los diálogos del Quijote. Análisis de la representación de las imágenes positivas y negativas de los protagonistas. *Oralia* 4: 129-148.
- Hutcheon, L. 1994. *Irony's Edge: The Theory and Politics of Irony*. London: Routledge.
- Kerbrat-Orecchioni, C. 1996. *La conversación*. Paris: Seuil.
- Kerbrat-Orecchioni, C. 2000. Gestion des conflits et constitution de coalitions dans les polylogues. En J. J. de Bustos Tovar; P. Charaudeau; J. L. Girón Alconchel; S. Iglesias Recuero y C. López Alonso coords. *Lengua, Discurso, Texto*, vol. 1. Madrid: Visor Libros, pp. 53-69.

- Koch, P. y Österreicher, W. 1985. Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch*, 36: 15-43.
- Lakoff, R. 1973. The logic of politeness: Or, Minding your p's and q's. *Papers from the Regional Meeting, Chicago Linguistic Society*, 9: 292-305.
- Lakoff, R. 1989. The limits of politeness. *Multilingua*, 8: 1001-1129.
- Lavandera, B. 1988. The social pragmatics of politeness forms. En U. Amnon y N. Dittma eds. *Sociolinguistics: An International Handbook of the Science of Language and Society*, 2. Berlin/New York: Walter De Gruyter, pp.1196-1204.
- Leech, G. 1983. *Principles of Pragmatics*. London: Longman.
- Marimón Llorca, C. 2008. *Análisis de textos en español. Teoría y práctica*. Alicante: Universidad de Alicante.
- Martín Rojo, L. 2000. Enfrentamiento y consenso en los debates parlamentarios sobre la política de inmigración en España. *Oralia*, 3: 113-148.
- Placencia, M. E. 2001. Percepciones y manifestaciones de la (des)cortesía en la atención al público, el caso de una institución pública ecuatoriana. *Oralia*, 4: 177-212.
- Raskin, V. 1985. *Semantic Mechanisms of Humor*. Dordrecht: Reidel.
- Rodríguez Marín, D. 2013. *Memecracia. Los virales que nos gobiernan*. Barcelona: Gestión 2000.
- Scolon, R. y Scolon, S. 1995. *Intercultural Communication. A Discourse Approach*. Oxford: Blackwell.
- Searle, J. 1986. *Actos de habla*. Madrid: Cátedra.

Sergio Rodríguez-Tapia (Universidad de Córdoba (España)) y Yordan Todorov Apostolov (Universidad de Córdoba (España)). Reseña de García, Dolores; Hernández, Héctor; Sinner, Carsten. 2019. *Clases y categorías en la semántica del español y sus interfaces*. Berlin (Alemania): De Gruyter

Clases y categorías en la semántica del español y sus interfaces ha sido editado por Dolores García Padrón, Héctor Hernández Arocha y Carsten Sinner. El volumen integra trabajos de perspectivas variadas que tienen por objetivo describir y explicar «los modos categorizadores del hecho lingüístico en la tradición lingüística hispánica» (p. 1), especialmente en relación con el funcionamiento del plano semántico. Así, en la obra que reseñamos es posible encontrar trabajos con un enfoque cognitivo, funcional o formal, entre otros, que prestan atención a lo sincrónico, a lo diacrónico y al contraste entre lenguas, y que exhiben, en su conjunto, la transversalidad con la que es posible investigar este objeto de trabajo. Esta obra se suma a las recientemente editadas por Sinner, que giran en torno a las clases y categorías de la lengua española, en colaboración con otros investigadores: Hernández Socas, Batista Rodríguez y Sinner (2018) sobre el contraste del español con otras lenguas; Adelstein, Bernal y Sinner (2019) sobre la formación de palabras del español; García Folgado, Sinner y Toscano y García (2019), que adopta una perspectiva historiográfica; o Sinner, Tabares Plasencia y Montoro del Arco (2020), en torno a la fraseología del español.

La obra se articula en torno a dos bloques temáticos con cinco capítulos cada uno: semántica y gramática, por una parte, y semántica y cognición, por otra. Estos diez capítulos son obra de trece investigadores que proceden de diez centros de investigación diferentes. Como indican los editores en la segunda página del volumen: «Los tópicos que reciben atención en los distintos capítulos de la monografía son, a grandes rasgos, la gramaticalización, la lexicalización, las clases léxicas, la composicionalidad, la estructura argumental, las equivalencias léxicas, la categoría gramatical nominal, la categoría gramatical verbal, especialmente el aspecto léxico, y las categorías de la lengua y la realidad». La organización de la obra atiende, principalmente, a la presentación previa de trabajos de corte general y, posteriormente, a los capítulos que se centran en aspectos específicos. No obstante, tal y como también avisan los editores, se trata simplemente de una forma de organizar los trabajos, puesto que en el primer bloque también es posible hallar perspectivas que abordan la cognición y, en el segundo, trabajos que atienden a los aspectos gramaticales, por mencionar solo algunas interrelaciones.

El primer bloque de la obra se inicia con «Sobre la naturaleza composicional de las unidades gramaticales», en el que Ignacio Bosque Muñoz toma la composicionalidad como objeto de análisis. Por un lado, Bosque estudia los problemas de la composicionalidad externa, los problemas de la interpretación semántica según su localidad (es decir, si la interpretación ocurre «paso a paso o de manera progresiva», p. 25) o la asignación de referencia en algunos definidos débiles. Por otro lado, Bosque explora los escollos en la composicionalidad interna, como la problemática clasificación de los pronombres *usted/ustedes* o la paradójica naturaleza de los pronombres reflexivos en el caso del reflexivo enfático en *Mi actuación habla de mí mismo* (p. 39), que incumple varias propiedades definitorias de los reflexivos, y que ya se avanzaron en

Bosque (2016). Cabe poner de relieve que el autor rechaza que la interpretación semántica se base en el mero análisis de argumentos aislados y toma en cuenta la información externa a la oración, especialmente en relación con la referencialidad (p. 40). Destaca también su postura a favor de la *granularidad* como estrategia analítica más adecuada, en contraposición con la *prototipicidad* o la *gradualidad* (pp. 15 o 40).

Dolores García Padrón y Gerd Wotjak, en su contribución titulada «¿Podría considerarse que el nombre propio es una categoría semiléxica?», estudian la naturaleza del nombre propio. Empiezan identificándolo lingüísticamente mediante ejemplos de antropónimos extraídos de la prensa española, como *La manada* o *Chiquito de la Calzada* (p. 52). A continuación, plantean el nombre propio como una entidad que reúne tanto contenido gramatical deíctico como contenido léxico (p. 58). Finalmente, se centran en la interacción de los procesos semánticos de propialización y despropialización en el neologismo *Las Kellys* (p. 62). De esta forma ejemplifican el potencial creativo de los mecanismos de la lengua para acuñar nombres propios nuevos. Los autores sostienen en su conclusión que el nombre propio constituye un universal semántico que posee función identificadora y simbólica (p. 66).

En la necesidad de mejorar la función codificadora de los diccionarios está centrado el capítulo «La categorización de las subclases léxicas en algunos diccionarios monolingües», de Rosario González Pérez, quien realiza un exhaustivo cotejo entre cuatro recopilaciones lexicográficas: *DLE*, *DUE*, *DEA* y *Clave*. Tras identificar las carencias de estos diccionarios en relación con la categorización léxico-semántica mediante hiperónimos de las unidades que definen (p. 90), la autora propone la adición de nuevas marcas para la identificación de las unidades léxicas: marca de sustantivo individual o colectivo, por ejemplo, en *árbol* y *arboleda* (p. 77), de aspectualidad, de gentilicio o de adjetivo de posibilidad, entre otras. Para concluir, destaca la conexión semántico-funcional de las palabras que la lexicografía debería abordar para poder potenciar su función codificadora.

La contribución «Desmontando la equivalencia. A propósito de equivalencias interlingüísticas en la semántica de verbos prefijados», de Elia Hernández Socas, se dedica a explorar las posibilidades de equivalencia entre varias lenguas. La autora compara la voz *ἀποικοδομέω* del griego clásico, *abbauen* del alemán, *destruir* y *desmontar* del español y *destruere* del latín, analizado detenidamente sus orígenes y su composición morfológica. Mediante numerosos ejemplos, cálculos lambda, glosas y plantillas léxicas, hace hincapié en la considerable dificultad de establecer una equivalencia plena entre vocablos: en alemán para hacer alusión a un despido se usa la colocación *Personal abbauen* (p. 107), no obstante, en español la colocación *desmontar el personal* (p. 111) sería errónea. En conclusión, pese a la aparente similitud formal entre las lenguas, existe un amplio abanico de restricciones de marcos contextuales y matices semánticos que convierten la equivalencia en un concepto gradual que debe analizarse en varios niveles (p. 127).

El siguiente trabajo, «Categorías de la lengua y categorías de la realidad», de Marcial Morera Pérez, se adentra en el histórico debate entre el pensamiento realista, según el cual la relación entre lengua y realidad discurre del referente a la palabra, y el pensamiento nominalista, que defiende una dirección palabra-referente. El autor sostiene la segunda postura aplicando cuatro planos de significación: primaria, categorial, morfológica y sintáctica a los ejemplos *cantante*, *destemplantar* y *fogalera*, y

explorando las orientaciones de sentido que estos adoptan en diversos contextos. Además, establece que las relaciones aleatorias (nos preguntamos aquí si esta expresión se refiere al concepto de arbitrariedad, como el autor lo emplea en la página 145) entre las palabras y las cosas pasan por una etapa subjetiva, que permite expresiones poéticas como *La mar con su canto eterno* (p. 147) y una etapa objetiva, que posibilita el uso convencionalizado y colectivo de los signos. Tras la argumentación contundente de su tesis, el investigador aboga por el replanteamiento de la tradición lingüística que reduce el constituyente semántico de las lenguas al mero significado de sus signos.

Juan Cuartero Otal inicia la segunda parte del libro con su estudio «Dificultades para la distinción entre predicados télicos y atélicos», que se enfoca en la acotación de los límites temporales de las acciones. Por medio de diversas pruebas, el autor demuestra que la distinción tradicional entre predicados télicos y predicados atélicos no es suficientemente precisa, sobre todo en el caso de los predicados de estado como *Se está pareciendo a su padre* (p. 161) y los eventos semelfactivos o puntuales como *Algo está explotando* (p. 163). En conclusión, el autor pone énfasis en la necesidad de un estudio más exhaustivo mediante un corpus vasto de ejemplos con el fin de solucionar la cuestión pendiente que gira en torno al análisis de la información tempoaspectual.

En el siguiente capítulo, «Un modelo onomasiológico y cognitivo para el análisis de la aspectualidad en las lenguas románicas», Sarah Dessì Schmid aborda la estructuración temporal de los estados de cosas partiendo de la interacción de las facultades perceptivas, categorizadoras y focalizadoras de la mente humana y la articulación *figura-fondo* (p. 183). En su modelo, la autora fusiona las tres dimensiones de la aspectualidad: externa, contextual e interna, en un enfoque monodimensional de tipo onomasiológico, pero controlado semasiológicamente (que recuerda al trabajo orientado a las lenguas romances publicado en 2019 por Dessì Schmid, también en De Gruyter), y se basa en la teoría de los marcos situacionales y el principio asociativo de contigüidad de Fillmore. Mediante la aplicación de este modelo a varios ejemplos, como *La vela se apagó* (p. 187), logra ilustrar de forma clara distintos casos de la compleja estructura aspectual del español.

En otro estudio dedicado a la aspectualidad titulado «La integración de la aspectualidad como categoría semántico-funcional en la lingüística española», Gerda Haßler y Verónica Böhm abordan una perspectiva interlingüística y diacrónica, comenzando por la reflexión teórica acerca de la aspectualidad en el griego clásico, el latín, las lenguas eslavas y el alemán. Posteriormente, analizan minuciosamente el desarrollo en la clasificación de los tiempos verbales del castellano en distintas gramáticas como las de Nebrija (1492), Bello (1847) o la *GRAE* de 1920, entre otras. En su conclusión, las autoras disciernen las lenguas aspectuales, en las que el núcleo de aspectualidad radica en el aspecto gramatical, de las lenguas romances, cuya aspectualidad se determina por los medios léxicos, los verbos incoativos o durativos, las perífrasis verbales o los medios contextuales (p. 212).

La contribución «El estado es eterno mientras dura», de Héctor Hernández Arocha y Guillermo Zecua, se centra en la estructura temporal-aspectual de los eventos (parten de las clases aspectuales de Vendler). En primer lugar, estudian la importancia del operador de cambio de estado *become*, poniendo especial énfasis en los problemas que plantea: las inferencias a las que da lugar, su gradación, como en el predicado culminativo *La puerta se abrió* (p. 226), y su opacidad en el caso de *El calor*

descongeló el hielo (p. 241). Teniendo en cuenta estos problemas, esbozan la implementación de un nuevo operador *state* y una perspectiva probabilística de la semántica de intervalos. Cabe subrayar que los autores logran ejemplificar e ilustrar mediante figuras su estudio para acercarlo a los lectores no avezados en estudios con perspectivas más formalistas.

El libro cierra con el trabajo «Expresiones de macroeventos en español: más allá de la tipología de lexicalización», de Yuko Morimoto, quien explica los macroeventos del español partiendo de la tipologización de Talmy entre las lenguas de marco verbal (como el español) y las lenguas de marco de satélite (como el inglés). La autora expone distintos casos en los que el inglés admite la estructuración eventiva correspondiente al marco verbal, como *He died from choking* o las posibilidades de uso del marco de satélite del español en *Se tiñó muy rubia* (p. 273). Por tanto, amplía la tipologización de Talmy, introduciendo el concepto de *macroevento*, y se basa en una óptica construccionista de la gramática con vistas a esclarecer la disparidad entre las construcciones con predicativo resultativo españolas y sus equivalentes del inglés. Este último estudio constituye un valioso paso en la investigación de la conceptualización cognitiva de la acción en las distintas lenguas.

Clases y categorías en la semántica del español y sus interfaces se caracteriza por plantear problemas de actualidad en la semántica hispánica que se presentan desde múltiples enfoques y en contraste con diferentes lenguas. La identificación de estos problemas viene acompañada de propuestas teóricas novedosas (como las planteadas en Morera o García Padrón y Wotjak, entre otros) y también metodológicas (véanse, por ejemplo, los trabajos de Bosque, Hernández Socas o Hernández Arocha y Zecua). Uno de los problemas identificados y a los que la obra presta una especial atención es la descripción y explicación del aspecto, para lo que se manejan perspectivas diferentes (cognitivas, formales, diacrónicas...). Tanto es así que más de la mitad de los capítulos abordan o mencionan algún problema relacionado con la aspectualidad (González Pérez, Hernández Socas, Cuartero Otal, Dessì Schmid, Haßler y Böhm, Hernández Arocha y Zecua, o Morimoto), lo que permite reflejar la complejidad del objeto de estudio y sus diversas manifestaciones en el plano léxico.

Con una edición cuidada y una organización que justifica la transversalidad con la que se puede abordar la semántica del español y sus interfaces, el volumen que hemos reseñado presenta trabajos de gran utilidad, tanto por el planteamiento de nuevos problemas, como por las propuestas teóricas y metodológicas que se esbozan como propuestas de solución.

Referencias bibliográficas

Adelstein, Andreína; Bernal, Elisenda; Sinner, Carsten, eds. 2019. *Clases y categorías en la formación de palabras del español*. Leipzig: Leipziger Universitätsverlag.

Bosque, Ignacio. 2016. La reflexividad paso a paso. En N. Rangponsumrit *et al.*, ed. *Monográficos SINOELE* 2:17, 15-58.

https://www.sinoele.org/images/Revista/17/monograficos/AAH_2016/AAH_2016_ignacio_bosque.pdf.

Dessì Schmid, Sarah. 2019. *Aspectuality: An Onomasiological Model Applied to the Romance Languages*. Berlin/Boston: De Gruyter.

García Folgado, María José; Sinner, Carsten; Toscano y García, Guillermo, eds. 2019. *Clases y categorías en la gramática española desde una perspectiva historiográfica*. Leipzig: Leipziger Universitätsverlag.

Hernández Socas, Elia; Batista Rodríguez, José Juan; Sinner, Carsten, eds. 2018. *Clases y categorías lingüísticas en contraste. Español y otras lenguas*. Berlin: Peter Lang.

Sinner, Carsten; Tabares Plasencia, Encarnación; Montoro del Arco, Esteban, eds. 2020. *Clases y categorías en la fraseología española*. Leipzig: Universitätsverlag.

Jesús Camacho Niño (Universidad de Jaén). Reseña de Contreras, Narciso M. 2020. *Estudios sobre el léxico del español. Diccionarios, variedades y enseñanza*. Jaén (España): UJA Editorial.

La obra *Estudios sobre el léxico del español. Diccionarios, variedades y enseñanza* recopila un conjunto de trabajos realizados por N. M. Contreras Izquierdo durante los últimos veinte años de su carrera investigadora. Todos los capítulos de la monografía giran en torno al componente léxico de la lengua española, si bien este se observa desde los múltiples enfoques que ofrecen ramas del conocimiento humano tan diversas como las lenguas de especialidad, la terminología, la variación lingüística o la enseñanza del español como lengua extranjera. Si bien el léxico constituye el eje central de la obra, hay un hilo conductor que subyace a toda ella, estructurándola y dotándola de coherencia y cohesión interna: el diccionario. Este se entiende como una herramienta didáctica que proporciona información lingüística y extralingüística sobre una selección de unidades léxicas, facilita el proceso de enseñanza y aprendizaje de lenguas y proporciona datos útiles para la investigación lingüística y léxica.

En cuanto a la estructura interna que presenta la obra, esta queda dividida en dos partes. El Bloque I –“Diccionarios, terminología y lenguas de especialidad”– consta de siete capítulos en los que se tratan aspectos relacionados con la presencia y el tratamiento de la terminología y las lenguas de especialidad en los diccionarios monolingües generales españoles en los siglos XIX, XX y XXI. A su vez, el Bloque II – “Diccionarios, variedades lingüísticas y enseñanza del español como lengua extranjera”– adopta un enfoque claramente didáctico y los tres capítulos que lo componen profundizan en el uso del diccionario como herramienta didáctica útil para la enseñanza de las variedades lingüísticas dentro del ámbito de ELE.

De forma más concreta, los capítulos uno y dos del Bloque I –“La lexicografía monolingüe del español en los siglos XIX y XX” y “*Ciencia y técnica* en la tradición lexicográfica española. Aproximación al tratamiento del conocimiento especializado en los diccionarios generales”– delimitan el marco epistemológico y el ámbito temático y temporal en el que se desarrollan el resto de capítulos de la sección.

El primero de ellos se acerca a las grandes empresas de la lexicografía monolingüe general del español, tanto académicas como no académicas, y analiza las principales novedades teóricas y técnicas que introducen estas obras. En el recorrido por la producción de la Academia, se abordan cuestiones relacionadas con aspectos tan diversos como los criterios empleados para la inclusión de voces de uso no general – tecnicismos, americanismos y neologismos–, la ordenación de la macroestructura –el uso de los dígrafos *ch* y *ll* como letras independientes–, la eliminación de las acepciones como artículos independientes o la reformulación de los sistemas de marcación diatómica. Sobre los diccionarios publicados fuera del seno de la institución, se hace especial hincapié en los elementos que empleaban los autores de estas obras para diferenciarlas de la académica y, por supuesto, superarla. De esta forma, se incide en aspectos como el aumento del lemario mediante la inclusión de regionalismos y tecnicismos, la introducción de información de carácter enciclopédico y sintáctico, la adopción de un enfoque más descriptivo que prescriptivo, la inclusión de elementos

paratextuales como las ilustraciones y los apéndices gramaticales, el uso sistemático de los ejemplos o la aplicación de la tecnología informática.

Por su parte, el capítulo dos realiza un recorrido histórico por el tratamiento que han recibido los conceptos *ciencia* y *técnica* en la lexicografía general española. Como punto de partida, el autor señala dos aspectos de importancia: la evolución del conocimiento científico y técnico se materializa en la lengua a través de la terminología y el tratamiento lexicográfico de estas unidades lingüísticas especializadas presenta problemas específicos como determinar qué términos deben registrarse en un diccionario general, cómo deben definirse para que sean comprensibles para un tipo de usuario potencial que no es experto y cómo deben estructurarse y clasificarse en el sistema de marcas especializadas. Esta presentación de los objetos de estudio facilita la aproximación a los conceptos de *ciencia* y *técnica*. En este sentido, se señala que la idea de ciencia no es un concepto fijo, sino cambiante y dependiente de factores históricos, culturales, sociales, etc. Esto permite al autor recorrer las distintas acepciones de *ciencia* que se suceden y conviven a lo largo de la historia y relacionarlas con el concepto de *técnica*, sobre el cual se concluye que, tradicionalmente, ha estado subordinado al de *ciencia*. De esta forma, la *técnica* se ha interpretado históricamente como un instrumento empleado por la *ciencia* para alcanzar sus objetivos y satisfacer sus necesidades. Sin embargo, señala que esta visión jerárquica cambia progresivamente desde la segunda mitad del siglo XX y especialmente desde la aparición del *Movimiento Ciencia, Tecnología y Sociedad*. Esta tendencia se caracteriza, entre otros aspectos, por adoptar una visión del conocimiento más social y cercana al ciudadano medio, rechazar ciertos tópicos como la primacía intelectual de las ciencias experimentales sobre las humanas y asumir el carácter interdisciplinar de la ciencia y la tecnología actual. A pesar de esto, en lo que respecta al tratamiento lexicográfico de estos conceptos, el autor apunta que, desde el siglo XVI, los diccionarios reflejan una visión aristotélica de la ciencia que se extiende hasta la actualidad, según la cual, esta constituye un “sistema ordenado de proposiciones derivadas de principios y causas” (Contreras Izquierdo, 2021, p. 13). Por tanto, estas herramientas de consulta no consideran otras visiones más actuales como la del *Movimiento Ciencia, Tecnología y Sociedad*. Esta postura tradicionalista se observa también en el tratamiento lexicográfico del concepto *técnica*. En este sentido, se advierte que no existe correspondencia entre la realidad lexicográfica y la social que deben representar las herramientas de consulta, lo cual tiene consecuencias que afectan a la práctica lexicográfica en cuatro aspectos relacionados con el tratamiento del léxico de especialidad: la selección de fuentes textuales, la inclusión de términos, su marcación y su definición. De esta forma, como conclusión final, remarca la necesidad de adaptar la visión del conocimiento especializado que transmiten los diccionarios a la situación social actual.

Los capítulos tres, cuatro y cinco –“El tratamiento del léxico especializado en la lexicografía española del siglo XIX”, “El léxico especializado en los diccionarios monolingües español del siglo XIX: el caso de la Física” y “Las definiciones de sustantivos técnicos en dos diccionarios del siglo XIX”– de este bloque adoptan un enfoque histórico e historiográfico y giran en torno al mismo tema: la presencia y el tratamiento del léxico de especialidad en los diccionarios monolingües generales del siglo XIX.

En el primero de ellos, se describe cómo ha evolucionado el tratamiento del léxico de especialidad en la lexicografía hispánica, con especial atención a la lexicografía

monolingüe general del siglo XIX, tanto académica como no académica. Para ello, el autor plantea una contextualización histórica, válida para este capítulo y los dos siguientes, sostenida en dos hechos: el nacimiento de la lexicografía de especialidad y la presencia del léxico especializado en los diccionarios generales. Sobre el primero, el autor expone que las voces de especialidad, inicialmente, se recogían en las nomenclaturas medievales y en los glosarios que se anexaban a tratados sobre ciencias y artes. Sin embargo, estos últimos, con el paso del tiempo y el avance científico y técnico, se hicieron cada vez más importantes, hasta el punto de que, en el siglo XVI, se independizaron, constituyendo herramientas de consulta autónomas. En cuanto al segundo hecho, el léxico de especialidad ha tenido cabida en los diccionarios monolingües generales desde su origen, tal y como puede observarse en obras como el *Tesoro de la lengua castellana o española* (1611) de S. de Covarrubias, el *Diccionario de autoridades* (1726-39) de la Real Academia Española o el *Diccionario con las voces de las ciencias y las artes* (1786-1793) de E. de Terreros.

A partir de esta contextualización histórica, el análisis del tratamiento lexicográfico de la terminología en los diccionarios generales españoles del siglo XIX toma como punto de partida la producción académica, concretamente, las innovaciones que se introducen en las obras publicadas durante esta centuria –1803-1899–, y para ello, parte de las referencias que se hacen al respecto en los prólogos de estas ediciones. Así, en la séptima edición (1832), destaca, por un lado, la apertura de la Academia a la inclusión de nuevas voces de especialidad que habían surgido gracias a los avances científicos que se estaban produciendo y, por otro lado, la reformulación de muchas entradas relacionadas con las ciencias naturales. En lo que respecta a las ediciones publicadas entre 1832 y 1843, se reformulan las definiciones de estas voces, buscando que estas no sean técnicas, ni excesivamente descriptivas e inexactas. En la novena edición del *DRAE* (1843), el prólogo de la obra se centra, principalmente, en defenderse de las críticas que recibía por parte de otros lexicógrafos y en justificar su postura y criterio de selección e inclusión de voces de especialidad: incluir solo las más generales. El tratamiento de este léxico no experimenta cambios significativos hasta la edición de 1884, donde se inicia una nueva etapa en su tratamiento. En este sentido, la décimo segunda edición supone un punto de inflexión en el quehacer lexicográfico de la Academia, que afecta también al tratamiento de las voces de especialidad, ya que la institución abandona su postura restrictiva y conservadora y abre la puerta a una inclusión más profusa de este tipo de voces. Sin embargo, como aduce el autor, esta apertura no pretende hacer de la obra académica un diccionario de carácter enciclopédico, pues mantiene su postura sobre esta cuestión: incluir solo las voces de especialidad más usuales. En lo que respecta a los diccionarios no académicos, en primer lugar, se exponen algunos factores que propiciaron la aparición de esta corriente lexicográfica: la pérdida parcial de la hegemonía lexicográfica que ostentó la Academia en buena parte del siglo XVIII, el aumento de las necesidades informativas de la sociedad y la competencia que existía en el mercado lexicográfico. La conjunción de estos aspectos provocó una ebullición lexicográfica que afectó no solo a la cantidad de diccionarios que se publicaron, sino también a los planteamientos y la técnica lexicográfica que se aplicaba a su construcción. Así, de entre los muchos diccionarios que se publicaron fuera de la Academia en el siglo XIX, el autor se centra en dos especialmente representativos de esta tendencia: el *Nuevo diccionario de la lengua castellana* (1846) de V. Salvá y el *Diccionario nacional* (1846-47) de R. J. Domínguez. Del primero de ellos, se dice que realiza una amplia revisión de la última edición del

DRAE que se había publicado –1843–, a la que incorpora numerosos términos perteneciente a ámbitos del conocimiento que, hasta el momento, apenas habían sido atendidos por la Academia, como la botánica. En cuanto a la obra de R. J. Domínguez, se apunta su dependencia de la misma edición del *DRAE* que empleó V. Salvá y se advierte como una de sus características más llamativas el enciclopedismo, que se plasma en un afán acumulativo que lleva a su autor a afirmar en el prólogo de la obra que esta registra más de ciento cincuenta mil voces de especialidad.

El segundo de estos capítulos transita el mismo camino que el anterior, si bien su desarrollo se acota al léxico perteneciente al campo de la Física. De esta forma, el autor limita su estudio a tres diccionarios: la décimo primera y décimo segunda edición del *DRAE* (1869 y 1884) y la décimo sexta edición del *Diccionario nacional* (1886). En cuanto a la metodología de análisis que aplica, esta se sostiene en tres elementos: la información que aportan los prólogos de las obras, las marcas de especialidad que incluyen y cómo se aplican, y un análisis comparativo de ambos sistemas de marcación. En lo que respecta a los prólogos, el autor destaca que el criterio seguido por la Academia es aumentar el número de voces de especialidad sin caer en la acumulación desmedida, mientras que el aplicado en el *Diccionario nacional* es acumulativo, porque se trata de una obra de carácter enciclopédico. El autor confirma esto mediante un recuento de las marcas de especialidad que recogen estos diccionarios: el *DRAE* (1869) registra 75 marcas de especialidad y el *DRAE* (1884), 73, mientras que el *Diccionario nacional* (1886) alcanza las 208. De forma más concreta, señala que en la décimo segunda edición del *DRAE* (1884), la Academia realiza una revisión y reformulación del sistema de marcas empleado en la edición anterior, cuyo resultado fue la supresión de un número importante de marcas de especialidad –albeitería, alfarería, ballestería, cabestería, cerería, cerrajería, cocina, dogma, drama, jineta, literatura, manejo de caballos, maquinaria, oratoria, platería, política y sestería– y la inclusión de otras –agrimensura, arqueología, bibliografía, equitación, fisiología, hidráulica, hidrometría, higiene, mecánica, minería, tauromaquia, tintorería, tipografía, trigonometría, veterinaria y zoología–. Por su parte, el *Diccionario nacional* incluye la mayoría de las marcas que se registran en la obra académica y muchas otras, las cuales son, en su mayoría, referentes a oficios –ballestería, banca, barbería, botonería, bordadores, calafatería, cantería, carbonería, carnicería, cerería, cordelería, cristalería, ebanistería, fontanería, jabonería, jardinería, sedería, tapicería, entre otras–.

En lo que respecta a la marcación de los términos propios de la Física, el autor observa, de un lado, que la edición de 1884 del *DRAE* aplica esta marca de especialidad –*Fís.*– a 25 entradas que no la poseían en la anterior, mientras que el *Diccionario nacional* (1886) la aplica a voces que “difícilmente pueden tener cabida en una obra como el *DRAE*, que pretende presentar solo el léxico general de la lengua” (Contreras Izquierdo, 2021, p. 33) y, de otro lado, que en la obra de R. J. Domínguez se marcan hasta 130 acepciones que no lo están en el diccionario académico, mientras que, al contrario, esto solo ocurre en 34 ocasiones. A partir de estos datos, el autor extrae varias conclusiones. En primer lugar, la edición de 1866 del *Diccionario nacional* no sigue un criterio claro y coherente en la asignación de las marcas de especialidad. En segundo lugar, el criterio de la Academia no es tan restrictivo como podría parecer *a priori*, pues el número de voces propias de la Física no se ha reducido, sino que muchas de las voces que se marcan en el *Diccionario nacional* no son consideradas tecnicismos por la Academia. En tercer lugar y relacionada con la anterior, el corpus más reducido que emplea la

Academia responde a que se trata de un diccionario general y no enciclopédico, como es el caso de la obra de R. J. Domínguez. En cuarto lugar y como conclusión final, la Academia adopta una postura más restrictiva y selecciona para su diccionario las voces técnicas que presentan mayor difusión entre los hablantes no especialistas, frente al *Diccionario nacional* (1886) que apuesta por la acumulación de datos, algo propio del enfoque enciclopedista que subyace a esta obra.

Finalmente, el tercero de estos tres capítulos toma como punto partida varios elementos que ya habían sido tratados previamente: el *DRAE* (1884), el *Diccionario nacional* (1886) de R. J. Domínguez y el léxico de especialidad propio de la Física, si bien, en esta ocasión, la atención se centra en el tipo de definición que se aplica a los sustantivos pertenecientes a esta rama del conocimiento humano. El punto de partida del autor es doble, por un lado, el concepto de definición lexicográfica como punto central de los diccionarios monolingües generales y, por otro lado, la diferenciación entre definición lingüística y definición enciclopédica. A partir de esto, se propone estudiar si la profusión en la inclusión de voces de especialidad va acompañada de una preocupación por su tratamiento definicional, y para ello, analiza cuatro aspectos: la corrección, la tipología, la información enciclopédica y la redacción. Sobre el primero, señala que las definiciones empleadas en ambos diccionarios cumplen, mayoritariamente, con la prueba de la sustitución y de identidad funcional, si bien, en algunas ocasiones, especialmente en lo que respecta a la obra de R. J. Domínguez, esto no pasa. Concretamente, señala que el *Diccionario nacional* (1886) emplea en sus definiciones palabras de la misma familia léxica del *definiendum*, así como artículos definidos al comienzo de estas, lo que genera problemas e incluso, llega a invalidarlas. Sobre la tipología definicional, ambos diccionarios optan por la definición sustancial con incluyente positivo para caracterizar este tipo de sustantivos, si bien el contenido de estas no hace referencia a sus características lingüísticas, sino extralingüísticas, por tanto, sostiene que no son definiciones propiamente dichas, sino descripciones del objeto designado por el término. Otra tipología definicional que está presente en ambos diccionarios es la de “falso incluyente”, es decir, aquella en la que “el género próximo no es independiente, sino que necesita el resto de la construcción para una correcta interpretación” (Contreras Izquierdo, 2021, p. 44). Sobre el tercer aspecto analizado, esto es, la cantidad de información enciclopédica que se incorpora a la definición de estas, el autor afirma que esta información debe gestionarse con precaución, ya que su fin es que el usuario sea capaz de identificar el objeto sin caer en la hiperespecificación, la cual no tiene cabida en un diccionario general y debe reservarse para los de especialidad. Esta reflexión sirve para introducir un tipo de definición que, desde la perspectiva del autor, es la más adecuada para el tratamiento lexicográfico de las voces de especialidad en los diccionarios generales: la *semienciclopédica*, la cual introduce datos enciclopédicos sin generar problemas de hiperespecificidad. En lo que respecta a la presencia de estas definiciones en las obras analizadas, en el *Diccionario nacional* abundan ambos tipos; sin embargo, en la obra académica, la mayoría son semienciclopédicas. Finalmente, el cuarto aspecto que tiene en cuenta hace referencia a la redacción de los textos definicionales y a la presencia de léxico de especialidad en ellos. En este sentido, el diccionario de R. J. Domínguez abusa del uso de estas voces, al igual que la Academia, si bien esta lo hace en menor medida, aunque en el prólogo advierte que ha intentado evitarlas. Todos estos aspectos llevan al autor a concluir el capítulo afirmando que la calidad de las definiciones del diccionario académico es mayor a las empleadas en la obra de R. J. Domínguez.

Finalmente, los últimos capítulos del Bloque I de la obra –“Terminología, variación lingüística y lexicografía: ¿una relación posible?” y “Procedimientos de creación de unidades léxicas especializadas”– adoptan un enfoque sincrónico y, en ellos, se abordan dos cuestiones fundamentales sobre las voces de especialidad: la interdisciplinariedad necesaria para su estudio y tratamiento lexicográfico, por un lado, y los mecanismos que posibilitan la generación de nuevos términos y la renovación de los existentes, por otro lado.

En el primero de estos capítulos, el autor parte del concepto de *variación lingüística* y cómo este varía en función de las características de los hablantes y de los recursos lingüísticos y extralingüísticos que despliegan estos según la situación comunicativa en la que se encuentren. A continuación, extrapola este concepto al ámbito de las lenguas de especialidad y sostiene que estas modalidades se ven afectadas igualmente, pues son parte de las lenguas naturales y como tales, hacen uso de todos sus recursos, si bien en distinta medida a la lengua general. Esto permite al autor señalar que el aspecto más característico de las lenguas de especialidad, frente a la lengua general, es el plano léxico, lo cual se conecta con otra disciplina: la terminología. A partir de aquí, el autor expone los fundamentos de la Teoría Comunicativa de la Terminología de E. Wüster. Esta defiende que la terminología es una rama autónoma del conocimiento humano que tiene como objetos de estudio el concepto y la noción, y cuya finalidad es fomentar la normalización terminológica y facilitar la comunicación científica internacional. Sin embargo, señala que su principal deficiencia es la concepción de los términos como elementos estáticos con “un carácter exclusivamente denominativo y estrictamente denotativo” (Contreras Izquierdo, 2021, p.54). Desde su punto de vista, estos planteamientos son insuficientes porque no tienen en cuenta la dimensión social de los términos y, por tanto, no pueden abordar la descripción del léxico y de la comunicación especializada. Frente a este modelo, se encuentra la Teoría Comunicativa de la Terminología desarrollada por M.^a T. Cabré. Esta propuesta adopta un enfoque lingüístico que entiende la terminología como un conjunto de unidades que pertenecen a una lengua natural, representan el conocimiento de una rama del saber y posibilitan la comunicación especializada. Con esto, el autor sostiene que la Teoría Comunicativa de la Terminología “defiende la variación del discurso especializado, asumiendo la diversificación discursiva en función de las distintas variables de la comunicación” (Contreras Izquierdo, 2021, p.55), lo que le permite corroborar la hipótesis que se presenta al comienzo del capítulo: las lenguas de especialidad se ven afectadas por la variación lingüística y, por tanto, constituyen una variedad más de las lenguas naturales.

A partir de estas reflexiones, el autor defiende la presencia de las lenguas de especialidad en los diccionarios generales, si bien señala que estas obras deben ofrecer, además de datos lingüísticos sobre estas unidades, datos pragmáticos relativos al ámbito temporal, geográfico, social, situacional y temático en el que se usan. De esta forma, el objetivo que persigue es analizar cómo se trata en los diccionarios generales la variación lingüística que presentan los términos de especialidad. Para ello, emplea un método analítico que parte de la selección de un conjunto de artículos lexicográficos pertenecientes a distintos diccionarios generales del español publicados a finales del siglo XX –*DRAE*, *CLAVE*, *DEA* y *DGLE*–. Esta metodología le permite extraer algunas conclusiones al respecto. En primer lugar, los diccionarios generales estudiados registran distintos lemas para designar un mismo concepto (sinonimia); en segundo lugar, incluyen términos con acepciones adscritas a distintos campos del saber, con

doble marcación y con referencias a voces del léxico general a las que se le han atribuido usos especializados en distintas disciplinas (polisemia); y, en tercer lugar, se registran términos que presentan marcas de transición semántica (usos figurados). Estos tres hechos van en contra de uno de los principios de la terminología clásica: la relación unívoca entre concepto y término, y según el autor, se explican por el trasvase que se produce entre la lengua general y la especializada, lo que denota que las unidades léxicas no pertenecen de manera exclusiva a una u otra, es decir, no son ni palabras, ni términos, sino elementos que tienen la capacidad de activarse de una u otra forma según el contexto y la situación en la que se usen. Como conclusión final, sostiene que el tratamiento lexicográfico de las voces de especialidad no ha evolucionado de forma paralela a los modelos teóricos de la terminología y, por tanto, los diccionarios generales ofrecen un enfoque más tradicional y cercano a la Teoría General de la Terminología, que a otros más actuales como la Teoría Comunicativa de la Terminología. Por último y sobre esta cuestión, apunta que, para paliar estas deficiencias, sería conveniente establecer puentes de trabajo entre terminólogos y lexicógrafos.

El último capítulo de este bloque es un trabajo que se inserta en el ámbito de la neología o neonomía. Es el único de toda la obra que no emplea el diccionario como eje vertebrador y su objetivo es presentar y analizar los distintos procesos que pone en juego el español para crear y renovar el léxico de especialidad. Con esto, el punto de partida y a la vez justificación es el aumento de la terminología especializada que se produce durante los siglos XX y XXI como consecuencia del desarrollo exponencial que ha experimentado el conocimiento especializado. Así, el autor explica que la lengua dispone de distintos recursos y procedimientos para dar respuesta a las necesidades denominativas que impone la actual sociedad del conocimiento. Estos procedimientos han sido descritos por numerosos especialistas, si bien el autor opta por la propuesta que hace Marcos Marín, pues “sintetiza de manera precisa los procedimientos de aparición o creación de un término” (Contreras Izquierdo, 2021, p.65). A partir de esta ella, analiza comparativamente los procedimientos más productivos con los que cuenta el español en el plano de la creación de unidades léxicas especializadas. En primer lugar, atiende a los procesos morfológicos de formación de palabras, concretamente, a la derivación y la composición. Sobre el primero, señala que recurre a morfemas y afijos que son conocidos y empleados en la lengua general y sobre el segundo, apunta que los formantes habitualmente son de origen grecolatino por el carácter internacional de las lenguas clásicas, si bien también hay casos de composiciones híbridas que combinan elementos compositivos cultos y patrimoniales. En segundo lugar, trata las léxias complejas, es decir, aquellas formadas a partir de unidades sintagmáticas, las cuales se rigen por dos reglas: coordinación y yuxtaposición, es decir, las mismas que se aplican en los sintagmas libres, de ahí que en ocasiones sean difíciles de identificar. En tercer lugar, hace referencia a distintos procesos de reducción léxica: siglas, abreviaturas, símbolos y acortamientos, sobre los que concluye que los primeros constituyen el proceso de acortamiento más productivo en el ámbito de la neología especializada. En cuarto y último lugar, se centra en el préstamo, el cual considera el proceso más productivo de todos los expuestos. Dentro de este, establece varios tipos: cultismos, préstamos y xenismos. Asimismo, señala que, aunque el préstamo sea el proceso más productivo, no está exento de ciertos problemas, los cuales pueden ser evitados adaptando el término prestado a las características de la lengua receptora, traduciéndolo u otorgando un nuevo valor a palabras ya existentes en la lengua. Finalmente, como

conclusión señala que es necesario establecer reglas y límites comunes que faciliten la creación de nuevos términos con el fin de mejorar la comunicación especializada.

Los capítulos que componen el Bloque II de la obra son tres –“El diccionario en la enseñanza del español como L2: el papel del usuario”, “La terminología en los diccionarios de aprendizaje del español” y “El diccionario, herramienta didáctica para la enseñanza y el aprendizaje de las variedades lingüísticas en el aula de español como lengua extranjera (ELE)”– y como se advertía al inicio, pertenecen al ámbito de la enseñanza de segundas lenguas, concretamente, a la del español como lengua extranjera. En ellos, se abordan cuestiones relacionados con el uso y aprovechamiento de los diccionarios como herramientas que facilitan la enseñanza y el aprendizaje de una L2.

En el primero, se analiza el papel que desempeña el usuario del diccionario dentro de un aula de español como L2. Para ello, el autor pone de manifiesto, en primer lugar, que los diccionarios constituyen una de las herramientas más útiles para los alumnos de L2, si bien en el ámbito de español como lengua extranjera, no se le ha prestado la atención que merece. En segundo lugar, señala que para obtener todo el potencial que ofrecen estas herramientas de consulta es completamente necesario establecer una tipología de sus usuarios potenciales y de las necesidades que presentan estos. Asimismo, también hace especial hincapié en que la formación lexicográfica de los docentes es imprescindible, pues son los encargados de guiar a los alumnos en la compleja tarea de elegir y consultar el diccionario que mejor se ajuste a sus necesidades. En este sentido, señala que los docentes tienen varias tareas al respecto, si bien la primera y más importante es presentar los tipos de diccionarios que tienen a su alcance –bilingüe en los niveles iniciales y monolingüe en los avanzados– y qué características tiene cada uno de ellos. A partir de estas premisas, el autor extrae conclusiones de los datos que arroja una encuesta sobre el uso de diccionario, la cual fue realizada por más de un centenar de alumnos de ELE pertenecientes a distintos centros de educación superior – Departamento de Estudios Hispánicos de la Universidad de JATE de Szeged (Hungría), Cátedra de Lingüística románica, italiana, española y portuguesa de la Universidad de Bucarest (Rumanía), Instituto Cervantes de Bucarest (Rumanía) y Cursos Internacionales de la Universidad de Salamanca–. Así, el autor plantea varias cuestiones en su encuesta:

- Frecuencia con la que se consulta el diccionario. La mayoría lo emplean muy a menudo.
- Tipo de diccionario que se consulta. La práctica totalidad de los encuestados afirma que usa diccionarios bilingües reversibles y muy pocos diccionarios monolingües.
- Motivos para elegir un diccionario. Una mínima parte de los encuestados afirma que lo eligen por su contenido, mientras que el resto aducen razones muy heterogéneas como recomendación del profesorado o el precio de la obra. Según el autor, esto evidencia la falta de preparación lexicográfica de los usuarios de estas obras.
- Qué diccionarios conocen. La mayoría no conoce otro que el que emplea habitualmente, si bien un reducido grupo se refiere al *DRAE*, al *DCECH* o al *DUE*.
- Qué diccionario español considera mejor. Un amplio número de estudiantes afirma no saberlo, mientras que un número reducido señala que no hay una obra mejor

que otra. Asimismo, hay alumnos que se guían por criterios como el que sea más grande o el que sea bilingüe.

- Qué formación ha recibido para usar el diccionario. Casi la totalidad afirma no haber recibido ningún tipo de instrucción lexicográfica y los que la han recibido, lo han hecho a través de sus padres, profesores o, incluso, bibliotecarios.

- Motivos que le llevan al uso del diccionario. En esta cuestión, los encuestados debían elegir entre dos opciones: interpretación o producción. En este sentido, la mayoría afirma que lo usan para entender los textos, si bien un número importante señala que lo usan en tareas de producción.

- Qué información consulta con más frecuencia. En esta cuestión, se daban varias opciones: semántica, gramatical, ortográfica, etimológica y enciclopédica. La mayoría contestó que habitualmente lo emplean para conocer el significado de una palabra, si bien hay otros que los usan para recabar información gramatical y ortográfica.

- Dificultades en el uso del diccionario. El principal problema con el que se encuentran los usuarios encuestados es no encontrar la palabra que buscan, seguido de problemas de interpretación de los textos lexicográficos.

Con estos datos, el autor concluye que los estudiantes consideran que el diccionario es una herramienta útil en su formación, especialmente, si este es bilingüe. Por tanto, es necesario diseñar y crear diccionarios que se ajusten a sus necesidades lexicográficas. Por último, estas herramientas deben incluir información gramatical y de uso, y, asimismo, deben presentarla de una forma clara y adaptada al usuario potencial de la obra.

El segundo capítulo de este bloque se acerca a una cuestión ampliamente tratada en el primer bloque de la obra: la presencia y tratamiento de la terminología y las lenguas de especialidad en las obras de consulta lexicográfica, si bien, a diferencia de los capítulos del Bloque I, en esta ocasión, la atención se dirige a los diccionarios de aprendizaje. El objetivo del estudio es reflexionar y obtener conclusiones en torno a si estas herramientas “deben incluir en su nomenclatura la terminología de los diversos campos de especialidad” (Contreras Izquierdo, 2021, p.85). Para ello, el autor parte de una selección de obras lexicográficas: *CLAVE*, *DIPELE*, *DUE* y *DSLE*, los cuales serán comparadas con el *DRAE*, y atiende a los siguientes aspectos: la información que presenta el prólogo sobre los criterios de selección de la terminología y el uso de las marcas de especialidad, el número de términos que incluyen y su distribución en distintas áreas temáticas, y la cantidad y uso de las marcas. En el caso de la primera cuestión, señala que el *DRAE* sí alude a los criterios de selección de términos: incluir aquellos que han sobrepasado su ámbito de especialidad y se han asentado en la lengua general. Frente a este, el *CLAVE* y el *DSLE* no aportan ningún tipo de información al respecto, mientras que, en el *DIPELE*, se afirma que, al tratarse de un diccionario de aprendizaje, no se incluyen muchos términos. Por último, el *DUE* presenta la misma nomenclatura que el *DRAE*, pero con algunas excepciones. En cuanto al uso de las marcas temáticas, este “no parece ser un aspecto esencial en las obras que hemos analizado” (Contreras Izquierdo, 2021, p.87), pues el *DRAE* no dice nada al respecto, mientras que en el *CLAVE* se afirma que se han incluido elementos que facilitan la localización de esta información. Más claro es el *DSLE*; en el prólogo de la obra, se afirma que emplea un sistema de marcación claro que ofrece información útil para usar las palabras en situaciones concretas. Finalmente, el *DUE* advierte que para marcar la

adscripción de un término a una rama del saber se emplean distintos tipos de acotaciones, si bien no se marcan los términos cuya adscripción especializada se evidencia en la definición o han pasado a la lengua general. Sobre el número de términos que se registran, el autor ha contabilizado todas las acepciones, dentro de la letra A, que incluyen marcas de especialidad. Como resultado, señala que el *DRAE*, con 2230, es el diccionario que más usos especializados registra, seguido del *DUE* con 724, el *DSLE* con 662, el *DIPELE* con 309 y, por último, el *CLAVE* con 258. En cuanto a la distribución por área de estos términos, apunta que las disciplinas que incluyen más voces son la *marina* –el *DRAE* incluye 292 y el *DUE*, 148–, la *medicina* –80 voces en el *DSLE* y 31 en el *CLAVE*– y finalmente, la *lingüística* –49 términos en el *DIPELE*–. Por último, sobre la aplicación de las marcas de especialidad, expone que los estudios sobre esta cuestión en el ámbito de la lexicografía de aprendizaje son escasos, si bien la mayoría de ellos coincide en que existe una falta de coherencia y uniformidad en su tratamiento y aplicación. Sobre las obras estudiadas, excepto el *CLAVE*, que no emplea ningún tipo de marca abreviada, el número de marcas de especialidad es muy dispar. Así, el *DRAE* incluye 153, el *DSLE*, 81, el *DIPELE*, 38 y el *DUE*, únicamente, 2. Según el autor, esta variedad y disparidad refleja la falta de un criterio común sobre la cantidad de marcas y su aplicación. Finalmente, como conclusión a su estudio, señala que la presencia del léxico de especialidad en los diccionarios de aprendizaje es una necesidad y, por tanto, todos los diccionarios las incluyen, si bien existen deficiencias que “pueden atribuirse a la carencia de unos criterios bien definidos para su tratamiento y, cómo no, a la dificultad intrínseca del objeto en cuestión, las terminologías especializadas” (Contreras Izquierdo, 2021, p.92). Entre estos problemas se señalan: la falta de indicaciones claras y precisas en los prólogos, la selección y el uso irregular de las marcas de especialidad y la desigualdad entre en el número de términos y su distribución por especialidades.

El último capítulo de este Bloque II es el más extenso de todos lo que componen la obra y se estructura en tres partes. La primera de ellas es de carácter teórico y en ella, se abordan cuestiones relacionadas con la variación lingüística y qué variedades de una lengua deben enseñarse en el aula de ELE. Sobre esta cuestión, el autor señala varias posturas al respecto: la primera es tomar como referencia una variedad prestigiosa de la lengua, la segunda es transmitir al alumno los elementos comunes a todas las variedades, es decir, enseñar la lengua general, y la tercera es analizar las necesidades de los alumnos y decidir qué variedad se enseñan en función de estas. A partir de estas tres posturas, concluye que lo más adecuado es partir de las necesidades de los alumnos y “que sea la norma culta de la zona donde el alumno estudia o con la que va a entrar en contacto por sus intereses la que se constituya en modelo de referencia” (Contreras Izquierdo, 2021, p. 96), ya que es la que presenta un mayor grado de homogeneidad. Sin embargo, también apunta que es necesario que los alumnos conozcan lo más característico de otras variedades. La segunda parte del capítulo es descriptiva y a lo largo de sus páginas, el autor presenta y analiza distintos recursos y herramientas que facilitan el proceso de aprendizaje de las variedades del español. De esta forma, pone de relieve la importancia del diccionario como herramienta fundamental para la formación lingüística y cultural de un hablante y, a partir de esto, presenta algunos recursos, la mayoría de naturaleza lexicográfica, que facilitan el aprendizaje de las variedades del español en el aula de ELE: *¿Usamos el diccionario?*; *Jergas de habla hispana*; *Así hablamos (Diccionario Latinoamericano)*; *Voces hispánicas*; *Hispanorama*, *Wikilengua*. *Atlas Oral*; *Panhispania Oral*; *Dialectoteca del Español*; *VARILEX*,

Variación Léxica del Español en el Mundo; Corpus de referencia del español actual (CREA), Corpus del Español del Siglo XXI (CORPES XXI); Corpus del español (M. Davies); Página de los diccionarios (J. R. Morala); Diccionario de variantes del español; Español coloquial; Diccionario del español coloquial; Español oral en contexto (C-Oral-Rom/ELE); Atlas interactivo de la entonación del español y Portal de lingüística hispánica. Finalmente, en la tercera parte del capítulo se presenta una batería de ejercicios dedicados a trabajar fenómenos pertenecientes a distintos niveles de análisis lingüístico del español a partir de los distintos recursos en línea que se presentaron en el apartado anterior. Concretamente, estos materiales están destinados a la enseñanza de las variedades dialectales en el nivel fonético-fonológico (seseo y ceceo), las variedades diastráticas en el nivel morfosintáctico (queísmo y dequeísmo), el español coloquial: nivel léxico y variación dialectal: nivel léxico.

Como conclusión, a lo largo de estas páginas, se ha puesto de manifiesto que la monografía publicada por N. M. Contreras Izquierdo, tanto en su planteamiento como en su desarrollo, ofrece una visión coherente y cohesionada sobre el léxico español visto desde las lenguas de especialidad, la terminología y la enseñanza del español como lengua extranjera, y a su vez, todo ello sustentado en el diccionario como herramienta fundamental para el estudio y la investigación lingüística en general y la enseñanza y aprendizaje de lenguas. Es una obra con evidentes fortalezas y muy útil para la investigación léxica y lexicográfica. En este sentido, el lector que se acerque a ella, encontrará datos de gran interés y relevancia para el estudio de la historia de lexicografía, el tratamiento lexicográfico de las voces de especialidad, la enseñanza de las variedades del español y el uso del diccionario en el ámbito de ELE.

Carmen Oliva Sanz (Universidad de Córdoba). Reseña de Guerrero, Gloria; Pérez, Manuel Fernando. 2020. *Terminología, Neología y Traducción*. Granada (España): Comares.

Han pasado más de treinta años desde que la Terminología pasó a formar parte de las asignaturas obligatorias de la licenciatura y del posterior grado en Traducción e Interpretación en España. Se trata de una disciplina muy relacionada tanto con la Neología como con la Traducción, disciplinas que se han desarrollado durante estos más de treinta años y han logrado establecerse como áreas del conocimiento totalmente afianzadas. Es por esto por lo que el presente libro, *Terminología, Neología y Traducción*, editado por Gloria Guerrero Ramos y Manuel Fernando Pérez Lagos, tiene como objetivo exponer el estado actual de las investigaciones que se están llevando a cabo en estas tres disciplinas, que en la mayoría de los casos se entrelazan para aportar datos y conclusiones interdisciplinarias.

Es cierto que los estudios sobre Terminología, Neología y Traducción no han dejado de desarrollarse y es posible encontrar numerosas obras que abordan las distintas perspectivas de estudio y las aplicaciones de estas tres materias. Es el caso de la compilación editada por Ruth Lavale-Ortiz en 2020, *Cognitivismo y neología: estudios teóricos y aplicados*; del libro *Traducción, Interpretación y Ciencia*, editado por José María Castellano Martínez y Aurora Ruiz Mezcuca en 2019, o del libro *Lenguas de especialidad, turismo y traducción*, coordinado por Miguel Ángel Candel Mora en 2021.

El presente libro aún no solo los últimos estudios de cada una de las disciplinas, sino que además incluye investigaciones interdisciplinarias en las que pueden apreciarse cómo la Terminología desempeña un papel fundamental para los estudios en Neología y cómo a su vez la Traducción resulta clave para la Neología. El enfoque del libro es tanto teórico como aplicado, de tal manera que muestra no solo las últimas reflexiones teóricas, sino también nuevas metodologías y recursos que pueden resultar de gran utilidad al futuro personal investigador y a los profesionales de estas disciplinas.

El libro está dividido en veinte capítulos que intercalan investigaciones de cada una de las disciplinas ya mencionadas. Cabe destacar que, aunque no hay un hilo conductor entre los capítulos y estos se pueden leer de manera independiente, la situación actual de crisis sanitaria derivada de la pandemia de COVID-19 ha influido en varias de las investigaciones, especialmente en lo que a aparición de neologismos se refiere. Además, tanto la traducción en el ámbito biosanitario como la terminología médica están muy presentes en el tomo. Muchas de las contribuciones, tal y como indican los propios autores, se escribieron durante el periodo de confinamiento de 2020. Por otro lado, se puede observar una cierta división del libro en dos grandes bloques, los primeros diez capítulos se centran más concretamente en reflexiones teóricas, mientras que la segunda mitad del libro presenta, en su mayoría estudios aplicados sobre las distintas disciplinas.

El primer capítulo, de Judit Freixa y Pedro J. Bueno «¿Son neologismos los ocasionalismos?» tiene como objetivo actualizar el concepto de neologismo para introducir, definir y clasificar los *ocasionalismos* y los *pseudoneologismos*. Destaca también la crítica a la utilización única del criterio lexicográfico para la detección de

candidatos a neologismos ya que en ese caso: «deben diferenciarse los *neologismos lexicográficos*, los candidatos que se obtienen con el criterio lexicográfico, de los *neologismos* propiamente dichos [...] que responden al mínimo exigido: carácter reciente y en proceso de institucionalización en el uso» (4). Sin duda, estas precisiones y nuevas clasificaciones resultan valiosas, pues el estudio de los *pseudoneologismos* y de los *ocasionalismos* muestra no solo la riqueza creativa de los hablantes, sino que también cumplen una función importante en el cambio lingüístico.

M.^a Teresa Cabré en «La teoría comunicativa de la terminología: fortalezas y debilidades» lleva a cabo una revisión de la TCT, una teoría que según la autora «está en permanente construcción» (21) y que evoluciona a medida que se precisa. El capítulo analiza tanto las principales fortalezas como las debilidades y puntos que han sido objeto de crítica de esta teoría. Esta última sección en la que se analizan y proponen soluciones a las debilidades de la TCT es, sin duda, la más novedosa del capítulo. Entre los aspectos comentados se encuentran la denominación del objeto, las categorías en cada plano del análisis del objeto y las categorías gramaticales y condición del término (26). Aunque a lo largo del capítulo se identifican también otras debilidades de la teoría, las limitaciones de espacio no permiten desarrollarlas y, en palabras de la autora, se ha optado por dejar «para otra ocasión los otros dos que ya hemos avanzado» (33). Sin duda el aspecto más destacable es la constante propuesta de soluciones, nuevas líneas y aspectos que pueden mejorarse para seguir construyendo una teoría integradora.

El tercer capítulo del libro es una contribución de Miguel Casas y se titula «La traducción como proceso siempre presente en el estudio interdisciplinar del lenguaje». A lo largo del capítulo se analiza el carácter lingüístico y no lingüístico de algunos de los problemas y dificultades de traducción, como es el caso de la traducción de sentidos y connotaciones inherentes a la realidad que llevan a una cierta «intraducibilidad de la realidad» (37). La interdisciplinariedad es un elemento clave en este capítulo. Mediante reflexiones teóricas desde distintas disciplinas lingüísticas y no lingüísticas como la antropología, la lógica y la filosofía, se trata de plasmar y dar explicación a fenómenos recurrentes y muy estudiados en las investigaciones lingüísticas.

Ana Belén Martínez y M. Gonzalo Claros dan, de nuevo, una llamada de atención sobre el desarrollo de nuestra lengua en su capítulo «Muchísimos neologismos científicos se acuñan en inglés: no dejemos su traducción en manos de los científicos», ya que junto con el desarrollo del inglés como lengua de referencia en ciencia y la precaria situación del desarrollo de la ciencia en España se preguntan «cuál es el futuro de una lengua que vive de espaldas a la lengua» (51). También analizan la falta de normalización en la traducción de tecnicismos y sus consecuencias, que oscilan entre las incongruencias y la creación de neologismos innecesarios y erróneos en español, todo ello ejemplificado con numerosos ejemplos muy conocidos. La contribución más novedosa del capítulo es el listado de fuentes fiables para profesionales del español que apuestan por una ciencia precisa terminológicamente (62-64).

El quinto capítulo corre a cargo de Mario Crespo y se titula «Creación semiautomática de una ontología terminológica médica a partir de FrameNet». A lo largo de la contribución no solo se establecen recursos informáticos para la gestión de terminología como son tesauros, corpus y ontologías, sino que también se expone un estudio en el que el investigador crea una ontología utilizando FrameNet mediante la semántica de los conceptos. Se trata de una línea de investigación que permite organizar el

conocimiento especializado «de la forma en la que los hablantes transmiten y entienden la información» (78-79), y los ejemplos que acompañan los distintos pasos de la metodología de este estudio resultan un apoyo muy práctico para la lectura y comprensión del proceso.

María Tadea Díaz se centra en el capítulo seis en «Los procedimientos morfológicos de la denominada neología formal». Realiza una nueva propuesta en la que se delimitan y diferencian los conceptos de neología formal, neología semántica y formaciones complejas. Tal y como indica la autora, hay que diferenciar las distintas formaciones y no todas ellas «pueden o deben ser consideradas candidatas a neologismo» (92). La originalidad y la propuesta de soluciones son la tónica del capítulo.

En la línea de capítulos previos, Rosa Estopà aborda la terminología médica en el séptimo capítulo «Terminología médica en sociedad: del uso especializado al diccionario de lengua». En su contribución, la investigadora analiza las características particulares de la terminología médica, como pueden ser el abuso de epónimos, los formantes clásicos y provenientes de lenguas como el francés o el alemán, o la prolífica variación terminológica. El trabajo hace énfasis en las dificultades comunicativas que plantea el discurso médico para el usuario lego que se enfrenta a informes médicos con terminología excesivamente compleja. Sin duda, la pandemia ha demostrado que es ahora cuando más se debe apostar por un lenguaje claro en medicina.

El capítulo octavo «La planificación terminológica del deporte como agente de política lingüística» corre a cargo de Francisco García, quien se centra en el ámbito deportivo y en las consecuencias de su planificación terminológica o la inexistencia de la misma. Se trata de un estudio comparativo entre países o regiones que han intervenido en la terminología del deporte de manera que han llevado a cabo una planificación implícita o explícita, y países o regiones que han desestimado esta intervención terminológica. Destaca, además, que los estudios lingüísticos previos relacionados con el deporte solo se centraban en el tecnolecto, mientras que la actual contribución estudia la planificación lingüística en lenguas mayoritarias como el francés y el español, en lenguas minoritarias como el bretón y el catalán, y en organizaciones multilingües como el Comité Olímpico Internacional y la UEFA. El enfoque novedoso de esta contribución y su apuesta por un lenguaje claro con numerosos ejemplos para demostrar las conclusiones del análisis resultan muy enriquecedores. Tras un exhaustivo análisis, el investigador concluye que «el deporte contiene una extraordinaria potencialidad como agente privilegiado de planificación lingüística implícita» (125) y que, además, «la acuñación de terminología propia equivale a una afirmación identitaria a través del lenguaje» (125).

Joaquín García explica en el capítulo octavo «Neología terminológica: retos y preguntas desde los nuevos caminos de la traducción» dos aspectos fundamentales para la terminología, la innovación y la internacionalización. Sin duda esta contribución es una clara llamada de atención hacia una problemática que ya se había estudiado con anterioridad: la terminología creada por y para el lingüista que deja de lado al usuario real de la misma. El investigador afirma que una terminología dirigida realmente al usuario «es capaz de abordar lo que este le demanda: que se encuentre en condiciones de poder ser usada en distintas situaciones comunicativas de carácter especializado» (133). Destaca especialmente una de las reflexiones del autor: «debería existir una teoría sobre la neología terminológica que [...] pudiera figurar dentro de una teoría general

sobre los neologismos» (135). Estas reflexiones sobre el estado de ambas disciplinas advierten sobre los riesgos de no cambiar el paradigma actual.

Gloria Guerrero y Manuel Fernando Pérez están a cargo del décimo capítulo «Terminología, Neología y Traducción: ¿es necesaria la traducción de neologismos?» donde comienzan con una serie de definiciones generales sobre las tres disciplinas mencionadas para, a continuación, centrarse en la traducción de neologismos. Si bien es cierto que afirman que «no se puede ni se debe traducir todo» (148), también reflexionan sobre el papel que merece el español en la ciencia. Siguiendo la línea de capítulos anteriores, estos dos investigadores abogan por una planificación para la traducción de neologismos en español que permita el avance del conocimiento especializado en nuestra lengua. A pesar de que las características y peculiaridades de esta necesaria planificación no se mencionan en detalle y una mayor ejemplificación podría enriquecer enormemente el capítulo, las reflexiones teóricas al respecto resultan de utilidad para futuras investigaciones y propuestas de aplicación.

La naturaleza y características lingüísticas del chino son el principal objetivo de la contribución de Yifan Li «La traducción conceptual» en el capítulo decimoprimer. Es por esto por lo que la investigadora afirma que «la traducción [...] muestra la carencia de interactuar entre las diferentes tipologías de las lenguas» (155) especialmente en lo que respecta a la lengua china. La propuesta de la investigadora de utilizar para esta lengua la traducción conceptual resulta peculiar, ya que este término parece hacer referencia a lo que actualmente conocemos por traducción funcional. Teniendo en cuenta la contribución de la autora es posible concluir que esta recién nombrada técnica de traducción es simplemente una herramienta específica para acercar la traducción moderna al idioma chino.

«Cuestiones gramaticales de los préstamos: neología y diccionarios» es el título del capítulo escrito por Mercè Lorente, en el que se presenta una investigación aplicada sobre la inclusión de neologismos en los diccionarios dependiendo de sus características. En este capítulo se pone de manifiesto la ya conocida resistencia de la RAE a incluir anglicismos naturalizados y la tendencia a su adaptación ortográfica. Además de clasificar los préstamos en distintas categorías, la investigadora analiza cómo las peculiaridades gramaticales de los anglicismos condicionan no solo su selección para incluirse en diccionarios sino su representación en ellos. En palabras de la autora: «la entrada de préstamos es natural en todas las lenguas» (181), pero la incorporación y adaptación de los mismos en diccionarios debe atender no solo a su frecuencia, sino también a todos sus aspectos lingüísticos entre los que no pueden olvidarse los elementos gramaticales.

El decimotercer trabajo corre a cargo de Nava Maroto y se titula «Cuando la informática lleva el timón: productividad y relevancia neológica del formante ciber- en el español actual». En su origen, el formante ciber- hacía referencia a robot o máquina, pero ha evolucionado y actualmente «se utiliza en la formación de neologismos relacionados con internet» (185). Además, se trata de un formante peculiar que se encuentra entre la prefijación y la composición. Lo más destacable de esta contribución es, sin duda, la utilización del baremo de relevancia neológica de Sánchez Ibáñez, mediante el que se pretende superar las limitaciones del criterio lexicográfico. Al aplicar este método, Nava Maroto llega a la conclusión de que el formante «lejos de desaparecer, está más vivo que nunca» (191), y su trabajo, en el que se detalla cada paso

de la metodología, se convierte casi en un manual para futuras investigaciones en neologismos.

M.^a Jesús Paredes dedica el decimocuarto capítulo a los «Usos terminológicos de la cata de vinos de Jerez», donde tiene como objetivo diferenciar entre lenguaje técnico y usos terminológicos en lo que respecta al caso concreto de los vinos de Jerez. Se trata de un léxico especializado «que posee ciertas particularidades propias que no comparte con otros lenguajes [...] situándolo en una controvertida frontera variacionista entre lenguaje común y lenguaje especializado» (199). La investigadora concluye que el lenguaje de la cata de vinos «presenta una importante problemática en su utilización tanto en consenso como en selección» (210) y que la terminología de los vinos de Jerez presenta una dificultad añadida debida a su caracterización técnica y a las consideraciones diatópicas. Son conclusiones muy concretas para un aspecto del discurso especializado muy particular y circunscrito a una región.

Maria Francisca Ribeiro de Raújo en «La traducción de los dobles en portugués y español por medio del análisis componencial» se centra en un problema al que todo traductor se enfrenta: la traducción de falsos amigos y las consecuencias que una mala traducción puede acarrear. Resulta muy relevante la metodología utilizada ya que se basa en un análisis componencial de los términos «que separa los rasgos que constituyen el significado de las palabras» (219) para evitar los errores en la traducción. Se trata de un análisis que muestra los rasgos presentes, ausentes o indiferentes en el significado de cada palabra para establecer cuál es la traducción óptima. Tal y como afirma la investigadora, este análisis «es un ejercicio mental que el traductor hace en un instante» (227), pero aplicar la técnica de manera consciente a palabras ambiguas, dobles o falsos amigos puede evitar errores de traducción. Sin duda esta herramienta puede ayudar a cualquier traductor independientemente de sus años de experiencia.

Miguel Sánchez en el decimosexto capítulo del libro, «Palabras en el punto de fuga: resistencias lexicográficas frente a la neología (o viceversa)», trata la controversia en la actualización de los lemas de los diccionarios. Las posturas de los hablantes y de la prensa frente al cambio lexicográfico en los diccionarios se han convertido casi en una ideología. No hay unos criterios absolutos ni un consenso en los criterios que debe cumplir una unidad para incluirse en un diccionario, pero tal y como indica el autor, todavía hay tiempo para que la sociedad conecte y acoja de manera ponderada y rigurosa la neología. Constituye una reflexión teórica esperanzadora que podría utilizarse como base para proponer una serie de criterios consensuados en futuras investigaciones. Se trata de una reflexión que pone de manifiesto cómo la lingüística no está exenta de ideologías.

Carmen Sánchez en «La traducción de neologismos semánticos por metáfora: variables y soluciones» explica cómo esta dificultad de traducción es un problema que debe tener en cuenta numerosas variables. La conceptualización de metáfora ha ido variando, pero su traducción siempre ha sido problemática. El análisis de las variables lingüísticas, pragmáticas y cognitivas que propone la autora ayuda a proponer soluciones para traducir neologismos semánticos por metáforas, aunque no se puede olvidar el carácter subjetivo y condicionado culturalmente de las mismas. El capítulo, aunque con reflexiones teóricas relevantes, se beneficiaría enormemente de ejemplos claros en los que se utilice el análisis propuesto para traducir neologismos concretos.

Isabel Santamaría lleva a cabo un análisis del tratamiento de los neologismos en un corpus compuesto por columnas periodísticas sobre la lengua. El capítulo, titulado «Entre la norma y el uso: los neologismos en las columnas sobre la lengua y el diccionario académico», demuestra cómo no solo los diccionarios no son objetivos y asépticos, sino cómo también las columnas sobre lengua construyen una imagen particular sobre la evolución lingüística. Aunque cada una de las tres columnas analizadas muestra una actitud frente a los anglicismos, la tendencia general sigue una actitud conservadora y purista, aunque tal y como expone la autora, las propias columnas hacen un uso importante de neologismos.

El penúltimo capítulo del libro corre a cargo de Carmen Varo y se titula «La neología como herramienta de análisis cognitivo de rasgos interlingüísticos: a propósito del discurso europeo generado en torno al coronavirus». Se trata de una investigación que aboga por la alianza entre las perspectivas cognitiva y social. Además, «aunque cada lengua explota en diverso grado los distintos mecanismos para la innovación léxica, todas confluyen en la incesante tendencia a la construcción de nuevas unidades» (279). Es esta tendencia la que genera un contagio entre las estructuras morfológicas entre neologismos de distintas lenguas. Esta contribución analiza además los procesos neológicos derivados de la actual pandemia lo cual resulta muy pertinente puesto que afecta a los distintos órdenes de nuestra vida y concluye que de todas las nuevas unidades «solo permanecerán en el acervo común [...] aquellas que trasciendan esa situación concreta a través de procesos semánticos de generalización» (291).

La última contribución del libro es la de Érika Vega, titulada «Las creaciones neológicas con elementos cultos en el lenguaje de la publicidad». Se trata de un estudio en el que la investigadora analiza los neologismos creados mediante procedimientos cultos que aparecen en la publicidad y concluye que estos formantes aportan científicidad, calidad y fiabilidad al producto o servicio anunciado. Las muestras reales utilizadas para el estudio nos permiten apreciar cómo la neología no solo está presente en prácticamente todos los aspectos de nuestra vida, sino que además influyen directamente a la hora de, por ejemplo, comprar cierto producto o contratar un servicio en particular.

Nos encontramos ante un extenso volumen que ofrece un acercamiento actual a las últimas investigaciones en Traducción, Neología y Terminología. Esta colección de investigaciones pone de manifiesto no solo la vitalidad de estas disciplinas, sino también su estabilidad como disciplinas perfectamente asentadas. Destaca la tendencia hacia el usuario que se mencionan en capítulos como el de Rosa Estopà y Joaquín García, y la relevancia de los estudios de Ana Belén Martínez y M. Gonzalo Claros, Mario Crespo y la ya mencionada Rosa Estopà para con la actual situación de pandemia sanitaria.

Resulta muy relevante, además, la variedad de aspectos tratados. En el caso de la neología, por ejemplo, las contribuciones no solo se centran en qué es un neologismo y cómo identificarlo, sino que también se definen nuevos conceptos como los de *ocasionalismos* y *pseudoneologismos* de Judit Freixa y Pedro J. Bueno. Lo mismo sucede en los trabajos sobre terminología, estos abarcan desde la planificación lingüística desde el deporte de Francisco García, hasta la terminología y usos terminológicos específicos de las catas de vino de Jerez en la contribución de M.^a Jesús Paredes, pasando por la utilización de nuevos recursos para crear ontologías en el

capítulo de Mario Crespo. Además, varias publicaciones como las de Érika Vega, Isabel Santamaría y Miguel Sánchez utilizan objetos de estudio realmente interesantes, novedosos y actuales como son las columnas de opinión sobre lengua o la publicidad. Sin duda esta perspectiva multidimensional tanto en los temas como en los enfoques que se tratan es uno de los puntos fuertes de este libro. En un solo tomo, el lector puede revisar los últimos estudios de las disciplinas de Traducción, Neología y Terminología.

Sí es cierto que la organización temática del tomo puede resultar algo compleja para el lector ya que no hay una agrupación clara de los temas. Aunque hay una ligera diferencia entre los diez primeros capítulos, basados en su mayoría en reflexiones teóricas, y los diez siguientes, en los que se ponen en práctica las distintas teorías lingüísticas, la diferenciación no es clara. Además, a nuestro parecer el lector se beneficiaría de ejemplificaciones sobre los nuevos aspectos teóricos y métodos propuestos en los distintos capítulos. A pesar de que las tres disciplinas sobre las que tratan los distintos capítulos están muy interrelacionadas, quizá una división del tomo en secciones dependiendo de la disciplina focalizada en el estudio facilitaría la lectura de los capítulos.

La modernidad del volumen y la numerosa variedad de temas y perspectivas de análisis son sin duda los puntos fuertes de esta publicación que forma parte de la colección Interlingua. Las reflexiones de los autores, que en ocasiones rescatan líneas de investigación planteadas en el pasado, no hacen sino reiterar la importancia de continuar la investigación para que el español adquiriera el papel que le corresponde dentro del mundo de la ciencia.

Referencias bibliográficas

Candel Mora, Miguel Ángel, ed. 2021. *Lenguas de especialidad, turismo y traducción*. Valencia: Tirant Lo Blanch.

Castellano Martínez, José María; Ruiz Mezcuca, Aurora. 2019. *Traducción, Interpretación y Ciencia*. Granada: Comares.

Lavale-Ortiz, R. M., ed. 2020. *Cognitivism y neología: Estudios teóricos y aplicados*. Madrid: Iberoamericana Editorial Vervuert.

Vicenta González Argüello (Universitat de Barcelona). Reseña de Lacorte, Manel; Reyes-Torres, Agustín. 2021. *Didáctica de la lengua como 2/L en el Siglo XXI*. Madrid: Arco Libros

La publicación de Manuel Lacorte y Agustín Reyes-Torres, *Didáctica del español como 2/L en el siglo XXI*, de la colección Manuales de formación de profesores de español 2/L, supone una revisión y actualización de los trabajos publicados anteriormente en relación con la formación del profesorado de español como segunda lengua y como lengua extranjera, como los de Herrera y Sans (eds.) 2015, Jiménez y Rufat (eds.) 2019, y Pujolà y López 2019 (eds.). En este nuevo volumen publicado por Arco/Libros, los autores, además de revisar los trabajos y teorías previos que nos permiten entender el estado en el que se encuentra en estos momentos la disciplina, ofrecen una puesta al día que tiene en cuenta las últimas novedades en los conceptos teóricos y las estrategias didácticas.

El libro, tal como afirman Lacorte y Reyes-Torres, va destinado, por una parte, a aquellos profesores experimentados que necesitan una actualización de sus conocimientos y de los procedimientos didácticos de mayor implantación. Asimismo, y puesto que se trata de una obra que tiene un claro objetivo formativo, otros posibles destinatarios son los que en estos momentos están formándose como profesores de español como segunda lengua o como lengua extranjera. Tanto unos como otros necesitan disponer de documentos actualizados en los que se recoja lo que se ha hecho hasta ahora en relación con la formación de profesores de español y también las nuevas propuestas que van surgiendo gracias a la consolidación de esta disciplina y la profesionalización de la enseñanza del español. Hay que añadir que por las características de la obra el público al que va destinado pueden ser profesores en ejercicio o futuros profesores de cualquier lengua extranjera y no solo de español. Prueba de esta afirmación son las referencias a documentos considerados clave en la didáctica de lenguas extranjeras, tanto en el contexto europeo (MCER, 2001; CFER, 2018) como en el contexto anglosajón (ACTFL, 2012; NSCB, 2016), y los elaborados por el Instituto Cervantes y el National Council of State Supervisors for Languages.

Los autores han dividido el volumen en 12 capítulos que dan cuenta de todos los aspectos que se consideran imprescindibles en la formación del profesorado. Estos van desde el primero centrado en la presentación de los conceptos fundamentales para una clara comprensión del resto de los capítulos hasta cuestiones como el enfoque que se sigue en la obra (el enfoque reflexivo), la planificación de clases, el desarrollo de las destrezas, la gestión del aula o la evaluación, como veremos con más detalle más adelante. Antes de iniciar la descripción del contenido de cada uno de los capítulos consideramos necesario detenernos en cómo se han articulado estos con el fin de entender mejor el carácter didáctico de esta obra.

Como ya se ha mencionado, lo que caracteriza esta obra es su claro interés didáctico, esa idea de formación del profesorado está ya presente en la estructura de cada uno de los capítulos. Así, cada capítulo se organiza en torno a cinco bloques. El primer bloque contiene unas preguntas de reflexión iniciales que intentan activar el marco de conocimientos previos que tiene el lector sobre el tema que se presenta. Como indican los autores de la obra, estas cuestiones preliminares pueden ser contestadas de forma

individual o pueden ser también compartidas con colegas de un mismo centro de trabajo o curso formativo. Todo ello con el fin de promover la reflexión sobre los temas tratados y, como consecuencia, la construcción conjunta de conocimiento a través de la posible interacción. Estas preguntas, a modo de andamiaje, ayudan al lector a que confirme lo ya conocido y a que pueda situar de forma ordenada los conocimientos nuevos que, gracias a la lectura, vaya incorporando, para alcanzar, así, una visión estructurada y organizada de la disciplina. Estas preguntas iniciales se considera que no tienen respuestas correctas o incorrectas, sino que persiguen el objetivo de promover la reflexión, además de dar la posibilidad al lector de incorporar su propia perspectiva ante el tema tratado.

Tras esas preguntas de reflexión iniciales, en cada uno de los capítulos se presenta un segundo bloque con los objetivos del mismo para, a continuación, centrarse en el tercer bloque, que contiene la información relevante del tema objeto de estudio. Este tercer bloque está dividido en diferentes subsecciones para facilitar su lectura. Además, en este tercer bloque, el lector encuentra referencias a unas preguntas de reflexión que cierran cada uno de los capítulos.

El cuarto bloque presenta una pequeña lista de lecturas recomendadas o bibliografía comentada sobre el tema. Estas lecturas seleccionadas por los autores tienen el objetivo de ofrecer una selección para que tanto el profesor en formación como el profesor experimentado que quiere continuar desarrollándose pueda ampliar los conocimientos tratados en ese capítulo.

El último bloque, como se ha mencionado, incorpora actividades finales de reflexión que persiguen el objetivo tanto de analizar los contenidos tratados en el capítulo como de facilitar al docente que ponga en relación esa información con su propio contexto laboral o su proceso formativo. Estas actividades invitan al lector a que contraste la información que se ha presentado a lo largo de cada uno de los capítulos con sus creencias o sus conocimientos previos y de esa forma ayudarle a construir conocimiento. Por todo ello, el lector de esta obra no puede adoptar una actitud pasiva, sino que se espera de él una lectura activa y reflexiva cuestionándose aquello que ya sabe, intentando incorporar nuevos conocimientos e intentando también que se cuestione algunas de sus prácticas didácticas a la luz de la información que se le proporciona y de ese modo ayudarle a poder actualizarse tanto en relación con conocimientos teóricos como con estrategias didácticas. Estas actividades de reflexión se pueden ir realizando a medida que se está leyendo el bloque tres, el de contenido, o dejarlas para el final, lo que le permite relacionarlas con las preguntas de reflexión iniciales del mismo capítulo. También hay que destacar las tablas y figuras que se incluyen a lo largo de la obra a modo de resumen, en las que se presentan las características más importantes de los aspectos que se están tratando en cada uno de los capítulos, con lo que se facilita enormemente la lectura.

Esta concepción de la obra y cómo se han articulado los diferentes capítulos está muy bien explicada por parte de los autores en el capítulo *Introducción*, en el que también se plantea a los lectores la libertad en cuanto al orden de lectura de los diferentes capítulos. La lectura del libro no necesariamente ha de seguir un orden lineal en cuanto al proceso de lectura, es una obra concebida para que el lector pueda empezar por aquel capítulo que sea de su interés o que pueda diseñar un recorrido en función de sus necesidades formativas o de actualización de conocimientos. En esta misma *Introducción*, en el

bloque central de contenidos, se sitúa el español en el mundo a partir de los documentos del Instituto Cervantes, para que el lector tome conciencia de la extensión geográfica de la lengua y de sus variedades. Además, se presentan las principales tendencias en la enseñanza de español como segunda lengua en contraste con las tendencias actuales. A continuación, se da paso a una revisión terminológica de los diferentes conceptos usados tanto en el ámbito de la enseñanza como de la investigación, y al español en el contexto de enseñanza cuando no es considerado lengua 1. Los autores también revisan los términos de *pedagogía*, *metodología* y *didáctica* poniéndolos en relación con *enfoque*, *método* y *técnica* en el contexto de enseñanza, usados a veces como sinónimos y que en realidad aluden a realidades diferentes. Este capítulo se cierra, como se ha mencionado más arriba, informando de la organización interna de cada uno de los capítulos, así como del uso de la obra como recurso en cursos de formación.

El segundo capítulo, *Tendencias actuales en la enseñanza comunicativa de 2L*, como su nombre indica, ofrece una presentación de lo que ha sido la enseñanza comunicativa desde sus inicios hasta la actualidad. Así se nos presentan enfoques de uso ya extendido como el Enfoque por tareas o el Aprendizaje basado en proyectos, pero también con otros de quizás no tanta implantación como la Enseñanza comunicativa intercultural o la Enseñanza comunicativa sociocultural. Estos dos nuevos enfoques nos llevan a plantearnos que la enseñanza de la lengua ha de ir más allá de la comunicación con el objetivo de facilitar la inclusión del estudiante en la comunidad de acogida de la lengua que esté aprendiendo. Para facilitar la comprensión de la Enseñanza comunicativa se ofrece una tabla resumen en la que se presentan los puntos centrales de la naturaleza del lenguaje y sus procesos de aprendizaje tal como se entienden desde esta perspectiva pedagógica, a la que le sigue un resumen de sus principales rasgos. A continuación, se presenta una revisión del Enfoque por tareas, teniendo en cuenta sus principios metodológicos. El tercer enfoque que se presenta en este capítulo es el Aprendizaje basado en proyectos, entendido como una continuación del Enfoque basado en tareas, pero que en la tradición española había tenido una mayor implantación en la etapa de educación primaria que en la enseñanza de lenguas extranjeras. Los proyectos se caracterizan por una mayor prolongación en el tiempo que las tareas, llegando a ocupar un trimestre o un curso completo. Se finaliza este apartado con la presentación de los criterios básicos que se han de considerar para llevarlo a cabo. El cuarto enfoque presentado es el de la Enseñanza comunicativa intercultural, perspectiva que dota a la cultura y a la competencia intercultural de un lugar central en la planificación de un curso de enseñanza de lengua. Así conceptos como la observación y participación en experiencias reales en las que se manifiesta la diversidad cultural, describir y analizar esa experiencia, contrastar experiencias y evaluarlas se consideran aspectos centrales de este enfoque. El último enfoque presentado es la Enseñanza comunicativa sociocultural, en la que se le da un valor preponderante a “la interacción social para el progreso de la cognición” (pág. 48). Entender la evolución del individuo solo es posible desde el contexto social y cultural en el que este está inmerso. Así, en este enfoque nos encontramos con nuevos conceptos apenas contemplados hasta el momento en la enseñanza como el de *Mediación* o el de *Interiorización* y con algunos ya considerados en otros enfoques como el de *Zona de desarrollo próximo* o *Andamiaje*.

Cierra este capítulo el apartado *Enseñanza de 2/L en el siglo XXI: Más allá de la comunicación*, en el que se nos presentan los conceptos clave que han ido configurando los enfoques de enseñanza hasta llegar al momento actual: la competencia

comunicativa, el papel central que adquiere el aprendiz de lenguas, los factores sociales que intervienen en el proceso de aprendizaje, la perspectiva multilingüe en el aprendizaje, la incorporación de teorías socioculturales y cognitivas, y la idea de contemplar las cuestiones de raza y racialización. Todo ello lleva a una consideración más dinámica, compleja y adaptable del lenguaje y a prestar más atención a las relaciones entre contexto social, relaciones de poder, identidad, emociones, ideología, etc. Desde esta perspectiva, el docente, en su proceso de desarrollo profesional, ha de adoptar una postura más reflexiva en relación con sus creencias y actitudes sobre los procesos de enseñanza y aprendizaje. Se termina el capítulo con la insistencia en dotar de más importancia al contexto, entendido como las condiciones personales de la comunidad que los participantes –de distintos grupos de edad, raza, género, origen, cultural o geográfico, etc.– crean en el aula de 2/L y que desde el primer día generan una serie de normas y rutinas respecto a las expectativas, responsabilidades y conductas individuales y colectivas. Todo ello sin olvidar las condiciones profesionales del docente y la necesidad de concienciarse del carácter político e ideológico presente en la toma de sus decisiones en torno a cuestiones como las variedades lingüísticas, las normas, los valores culturales, etc.

El capítulo 3, *La enseñanza reflexiva de 2/L*, persigue fomentar la práctica de la enseñanza reflexiva. Para ello el docente ha de llevar a cabo un trabajo de introspección en el que tome conciencia de sus creencias en relación con la enseñanza, revise sus experiencias previas de aprendizaje, reflexione sobre cómo se aprende y enseña, sobre su práctica docente, sobre sus planificaciones, sobre las decisiones que toma, sobre lo que le funciona y lo que no, sobre el tipo de profesor que se ha propuesto ser, sobre el trabajo que realizan los alumnos, etc. A partir de este trabajo de reflexión es como el profesor se convierte en un *Docente reflexivo* capaz de establecer su forma de enseñar, mejorar y ser más flexible. El profesor ha de reconocer su propio estilo de enseñanza, y cuál es el enfoque más efectivo y que mejor se adapta a sus estudiantes y a su contexto de enseñanza. Para facilitar la comprensión de ese modelo docente, los autores nos presentan criterios de reflexión, así como un modelo de práctica del docente reflexivo. Continúa el capítulo con la propuesta por parte de los autores de 10 principios educativos, basados en los resultados del proyecto TLRP (pág. 68), que pueden ser útiles para la toma de decisiones profesionales por parte de profesores. Esta enseñanza reflexiva dota a la interacción entre docentes y estudiantes de un papel central en el proceso de aprendizaje. Dicha interacción ha de entenderse no solo como el enriquecimiento de los estudiantes gracias a su relación con el profesor, sino también como un proceso de aprendizaje mutuo y de gran valor en el desarrollo profesional del docente. Como apuntan los autores, el resultado de la reflexión crítica y sistemática sobre los procesos de enseñanza aprendizaje del profesor ha de llevarle a establecer una filosofía personal de enseñanza. Esta filosofía le ha de guiar en la construcción de su identidad docente a partir de las relaciones que establezca entre su formación inicial, sus propias ideas o experiencias previas como aprendiz y sus experiencias como docente.

Los autores, en el cuarto capítulo, *Planificación, secuenciación y contextualización*, nos presentan las diferencias entre los conceptos clave en relación con la planificación: currículo, programa y lección, y los documentos, tanto europeos como americanos, que se han de tener en cuenta para poder llevar a cabo la enseñanza de lenguas extranjeras atendiendo a las necesidades de los alumnos, su nivel y su contexto de aprendizaje. La perspectiva que se adopta en relación al currículo es la del “*Can-Do*”, entendida como

la capacidad que tiene el aprendiz de identificar, establecer y evaluar sus propios objetivos de aprendizaje, y dar apoyo a la planificación de currículos y cursos con el objetivo de allanar la valoración de los diferentes niveles de comunicación en la lengua meta. Para los cinco componentes del currículo (componente gramatical, pragmático-discursivo, nocional, cultural y de aprendizaje) se basa en el Plan curricular del Instituto Cervantes. Los autores abordan la planificación de clases incorporando al alumno también en ese proceso de planificación de las sesiones de enseñanza. Se aboga en la obra por un papel activo del aprendiz de segundas lenguas, al que hay que escuchar y atender y con el que se puede negociar qué planificación se llevará a cabo. También nos ayuda este cuarto capítulo con el diseño de objetivos de programas de enseñanza de español y de planificación de unidades didácticas. En este sentido es muy clarificador el punto 4.6 *Planificación de clases y secuenciación didáctica* al plantear diferentes enfoques a la hora de llevar a cabo una secuenciación, apartado que muestra una total coherencia con el resto de la obra. Si bien, como se había dicho en el capítulo 2, cuando se nos habla de las tendencias actuales, se incorporan no solo las tareas, los proyectos, sino también la enseñanza desde una perspectiva comunicativa intercultural y desde una perspectiva sociocultural, en este capítulo al describir el proceso sobre secuenciación didáctica se retoman todos los enfoques mencionados. De este modo, se presenta cómo secuenciar si se sigue el enfoque por tareas, cómo hacerlo en el aprendizaje basado en proyectos, cómo en la enseñanza comunicativa intercultural y cómo en la enseñanza comunicativa sociocultural. Todo ello contribuye a que el lector pueda concretar la información de los capítulos previos.

El quinto capítulo, *La pedagogía de las multiliteracidades: un enfoque para la enseñanza del español 2/L en el siglo 21*, aporta una perspectiva de actualización de contenidos explícita en relación con este nuevo concepto y la necesidad de revisión por parte del profesorado. Adoptar este nuevo enfoque pasa por entender la necesidad de ampliar los objetivos de enseñanza, pasar de la mera comprensión de las posibilidades de comunicación, ampliadas considerablemente por el acceso a la red, a “incorporar una pedagogía que capacite a los alumnos para leer, ver, interpretar, hablar, escribir y reflexionar de modo crítico” (pág. 110). Se entiende así que la literacidad no se circunscribe solo al ámbito de enseñanza, sino que se amplía al contexto en el que está inmerso el estudiante para que este sea capaz de interpretar las prácticas sociales culturales e históricas y a la vez crear también significados. Esta forma de contemplar la enseñanza implica considerar las dimensiones personales y cognitivas (el deseo de aprender del estudiante), conceptuales (dominio de la lengua, la cultura, las manifestaciones artísticas, etc.), socioculturales y estéticas (contexto sociocultural). Así, desde la perspectiva pedagógica, este nuevo enfoque implica concebir a los alumnos como individuos capaces de producir contenido de múltiples modos y para contextos diversos y no solo de consumirlo. Y al docente se le concibe como un mediador multimodal o diseñador de experiencias de aprendizaje.

El capítulo sexto, *Análisis, diseño e implementación de recursos didácticos*, nos proporciona una distinción conceptual entre *manual* o *libro*, *materiales* y *recursos*, entendiendo que, en el momento actual, con la extensión del uso de la red, el material para la enseñanza se ha ampliado considerablemente y la mayoría de los profesores no se limitan en el aula al uso de libros o recursos analógicos. Los autores destacan el papel predominante de los libros en la enseñanza debido a su función como eje vertebrador del currículo para cursos de todo tipo, la ayuda que ofrecen al profesorado para

responder a las necesidades de aprendizaje, por su papel formativo entre el profesorado novel o con poca experiencia, etc. Todo ello sin perder de vista la escasa relación que a veces existe entre la investigación en el ámbito de español 2/L y la enseñanza, los pocos recursos con que se cuenta en el ámbito editorial para las publicaciones de este tipo o la incorporación masiva de las nuevas tecnologías a la educación (plataformas digitales de enseñanza, espacios web institucionales o personales, etc.), que hace que muchos profesores se estén distanciando del libro de texto para diseñar sus propios materiales de enseñanza.

El contenido central de este capítulo gira en torno al valor del texto en los procesos de enseñanza y los criterios de análisis de materiales. En relación con el primer aspecto se ha de destacar el valor preponderante que ocupan los textos de todo tipo en los actuales enfoques de enseñanza: se ha pasado del texto como pretexto a situarlo como principal recurso pedagógico, a permitir que a través de él los estudiantes puedan explorar otras dimensiones culturales y establezcan relaciones con sus propios conocimientos y experiencias y no limitarse solo a poner el foco en la lengua. Se cierra el capítulo con dos tablas que recogen pautas para el análisis y evaluación de libros y materiales, pero que también pueden ser de gran utilidad para los procesos de adaptación de los materiales existentes a grupos de alumnos y contextos específicos.

Los capítulos 7 y 8 son los centrados en las destrezas. Si bien los autores, al inicio del capítulo 7, *Desarrollo de las destrezas comunicativas: escuchar, ver leer*, manifiestan que la tendencia actual es la de un tratamiento integrado de las destrezas en el aula, mantienen la división tradicional con el fin de lograr una mayor profundidad en el tratamiento de cada una de ellas. En este capítulo junto con el 8, *Desarrollo de las destrezas comunicativas: escribir, hablar y pronunciar*, a partir de los documentos de referencia para establecer los marcos curriculares en la enseñanza de 2/L en EE.UU. y en Europa, nos presentan cómo se conciben y se integran las destrezas en ambos contextos. Estos se caracterizan por poner el énfasis en el valor comunicativo y social del texto y no solo en los contenidos y las destrezas de modo aislado. A partir de esa información se presentan las destrezas de forma aislada, haciendo especial hincapié en los aspectos pedagógicos y consideraciones didácticas para su tratamiento en el aula. En el capítulo 8, merece la pena mencionar los apartados dedicados a la *Conversación pedagógica*, sus características y sugerencias para su tratamiento en el aula; al *Tratamiento de errores orales*, en el que se nos informa de las posibles causas de los errores y cómo aproximarnos a ellos en el aula; y el *¿Y qué hacemos con la pronunciación?*, en el que se nos informa de sus objetivos, contenidos, materiales y nos ofrecen una tipología de actividades.

El capítulo 9, *Enseñanza y aprendizaje de gramática y vocabulario*, en el bloque inicial de contenido, nos aporta información sobre el tratamiento histórico de la gramática en el aula, sobre la importancia que en estos momentos se le da a la adquisición y enseñanza del léxico para, a continuación, tratar ambos aspectos por separado. En relación con la gramática se nos avisa de la contradicción existente entre el valor que se le otorga a la gramática pedagógica en estos momentos y cómo se sigue trabajando la gramática en el aula. Para profundizar en este contraste aportan información sobre las diferencias entre la enseñanza implícita de la gramática y la enseñanza explícita, y sobre el valor de los textos para la enseñanza de la gramática para así poder presentarla en su contexto natural. En relación con la didáctica de la gramática se presentan el *Enfoque en la forma* y el *Enfoque conceptual*, como dos perspectivas de aproximación a la misma.

El apartado centrado en el vocabulario aporta información más que suficiente para entender cómo se concibe este en la actualidad, incorporando no solo las unidades léxicas monoverbales, sino también colocaciones y fórmulas de interacción social, todo ello gracias a lo que se sabe en estos momentos sobre adquisición y a las investigaciones centradas en el análisis de corpus lingüísticos. Se cierra este bloque con consideraciones sobre técnicas específicas para la enseñanza de unidades léxicas.

Los autores, en el capítulo 10, *Lectura y educación literaria en 2/L: Una aproximación multimodal*, hacen una apuesta clara y valiente por recuperar los textos literarios en la enseñanza de 2/L. La novedad estriba en el concepto *Educación literaria* y su aplicación a la pedagogía de las multiliteracidades. Así, los autores proponen que el objetivo que se ha de perseguir en el aula es promover un aprendizaje que fusione “la participación subjetiva y emocional del lector con el dominio de ciertas convenciones para desarrollar el pensamiento crítico” (pág. 240) y no limitarse a presentar la historia de la literatura o la transmisión de contenidos a través de la lectura. Esto tiene incidencias en la formación del profesorado al esperar que el docente sea capaz de llevar al aula todo tipo de lecturas (novelas, poemas, relatos, álbumes ilustrados, etc.) a partir de las cuales el estudiante pueda llegar a establecer un proceso de transacción de significados a partir de una comprensión que integre texto e imágenes. El profesor, desde este enfoque, se convierte en un mediador en el análisis de textos e imágenes. Para conseguir ese objetivo el docente ha de ser capaz de potenciar que los alumnos partan de su identidad para la lectura interpretativa, facilitar el acceso a textos y temas que sean de su interés, ofrecer tiempo para la lectura y guiarles en el proceso de aprender la forma de crear significado a partir de textos literarios y multimodales. Continúan los autores ofreciendo un marco para el desarrollo de la competencia literaria, para acabar el capítulo presentando un modelo de análisis para guiar al profesorado en la implementación de la pedagogía de las multiliteracidades a través de un álbum ilustrado.

El capítulo 11, *La clase de 2/L como aspecto social: Afecto, gestión y liderazgo*, se centra en cuestiones que no siempre son tenidas en cuenta en la formación del profesorado y que son decisivas para conseguir un ambiente propicio hacia el aprendizaje. Entender el espacio del aula como un microcosmos en el que cada día confluyen alumnos y docente para compartir objetivos complementarios ayuda a entender la importancia de las rutinas de clase, de la cultura que se genera entre el grupo hasta llegar a considerarla como una comunidad de práctica. Es desde esta perspectiva que cobra especial relevancia la dimensión afectiva entendiendo que si entre el grupo hay un buen ambiente los estudiantes se sentirán más predispuestos a aprender. Por ello es importante que el docente establezca relaciones de proximidad entre todos los miembros del grupo clase, que sepa establecer un equilibrio entre el afecto y la autoridad. A continuación, los autores presentan rutinas y estrategias que se han de considerar para una óptima gestión del aula, como por ejemplo las relacionadas con los inicios de clase, los finales, la planificación e implementación de las lecciones y el conocimiento del grupo. Todo ello sin olvidar la importancia de dotar de autonomía al alumnado y de un uso del lenguaje apropiado para salvaguardar la buena disciplina, dinámicas que favorezcan un buen ritmo y máxima interacción. Se cierra el capítulo con un apartado dedicado al liderazgo docente en tanto que este puede ejercer una gran influencia en el proceso de aprendizaje.

Termina la obra con el capítulo 12, *Evaluación*. Se contempla este aspecto de la enseñanza como una parte más de la educación integral del estudiante. Se parte de una perspectiva coherente con el enfoque de formación que proponen los autores: “Si el objetivo del aprendizaje es desarrollar la literacidad para construir significados, interpretar diferentes tipos de textos y conceptos, comunicarse y expresar sus ideas, el docente ha de marcarse como meta evaluar la capacidad de sus alumnos para reflexionar y desarrollar nuevos conocimientos en la 2/L” (pág. 290). Los autores hacen una breve revisión histórica sobre la evaluación en el ámbito de la 2/L para detenerse en la evaluación desde la perspectiva de la pedagogía de las multiliteracidades. Así, desde esta perspectiva, la evaluación ha de estar contextualizada, ser multidimensional y tener un carácter pedagógico. El siguiente concepto que se presenta es el de *retroalimentación*, diferenciando lo que es el tratamiento del error durante la instrucción (en el transcurso de las sesiones de clase) de la retroalimentación dada a actividades ya completadas para orientar al estudiante en su aprendizaje posterior.

La segunda parte de este capítulo 12 se centra en las herramientas de evaluación, para las que se plantean recomendaciones de diseño y administración: pruebas y exámenes, rúbricas, evaluación integrada de la actuación (sería la aplicada a proyectos, por ejemplo, en la que cada una de las fases incluye algún tipo de retroalimentación o rúbrica) y portafolios. Concluye este apartado con ideas sobre cómo incorporar la autoevaluación y la evaluación entre iguales en el aula.

Además de las lecturas recomendadas que acompañan cada uno de los capítulos el libro, el volumen se cierra con un extenso apartado bibliográfico en el que se recogen todas las referencias mencionadas a lo largo de la obra: un buen punto de partida para aquellos profesores en formación o en ejercicio que tras leer la obra deseen continuar su desarrollo profesional.

Referencias bibliográficas

Herrera, Francisco; Sans, Neus, eds. 2015. *La formación del profesorado de español. Innovación y reto*. Barcelona: Difusión S.L.

Jiménez Calderón, Francisco; Rufat, Ana, eds. 2019. *Manual de formación para profesoras de ELE*. Madrid: SGEL.

Pujolà, Joan Tomàs; López, Carmen, eds. 2019. *La didáctica de lenguas de par en par. Diálogo entre teoría y práctica. Homenaje a Ernesto Martín Paris*. Barcelona: Difusión S.L.

Álvaro Molina García (Universidad de Málaga). Reseña de *Iberoromania*. *Contacto y migración. Desafíos metodológicos en la sociolingüística hispánica actual* 91 (2020).

El número 91 de la Revista *Iberoromania* abarca un tema vigente y de gran importancia: situaciones de contacto de lengua y migración. La mayor facilidad de los hablantes para desplazarse, las conexiones comunicativas a diario a nivel internacional y, en definitiva, la gran cantidad de exposición a diversas lenguas a las que los hablantes se ven confrontados han provocado, efectivamente, que los cimientos de los estudios de contacto de lengua se tambaleen y se requieran nuevos cauces de investigación. Un total de siete artículos reflejan diferentes puntos de vista sobre la cuestión, poniendo el foco principalmente en dos elementos: por un lado, en una reflexión metodológica y en la importancia de considerar una combinación de métodos cualitativos y cuantitativos y, por otro lado, en el debate sobre la posible necesidad de desechar la antigua idea de las lenguas como entidades monolíticas y fijas y de adoptar, por el contrario, la concepción del lenguaje como una suma de recursos lingüísticos con significados sociales asociados a distintas variedades lingüísticas a disposición de los hablantes.

Yvette Bürki y Caroline Patzelt (primer artículo) comienzan con una síntesis y una estructuración del monográfico. En los últimos años, las investigaciones sobre el español han vivido un auge de estudios cuantitativos de la mano de Villena Ponsoda y Moreno Fernández. Sin embargo, no debe caer en el olvido la necesidad de recuperar, en cierta medida, los factores cualitativos de corte sociopragmático, que permiten conocer la influencia de las identidades y los factores socioculturales. Asimismo, expresan la necesidad de ampliar el conocimiento sobre los procesos de renegociación de posturas sociales mediante la actuación lingüística en los estudios de migración y contacto de lenguas. De esta manera, los hablantes se valen de formas lingüísticas a las que se les asocia un determinado significado social que se va renegociando durante el transcurso del acto comunicativo, lo que da lugar a los procesos de acomodación o de divergencia; pero no se trata de una elección dicotómica entre variedad A o variedad B, sino que el posicionamiento se forma a partir de una compleja microelección de recursos lingüísticos procedentes de diferentes variedades o lenguas en función de la consideración que de estas tienen los hablantes migrantes.

Adriana Patiño-Santos propone una aproximación etnográfica a las narrativas en los contextos de migración basada en la noción de reflexividad (segundo artículo). En primer lugar, presenta un marco teórico y un estado de la cuestión bastante sólido y ampliamente documentado, asentándose sobre dos conceptos clave: las narrativas y los tipos de reflexividad. La autora justifica la propuesta de esta aproximación etnográfica por la desestabilización de la figura del Estado como eje identitario de los individuos (Codó, Patiño-Santos y Unamuno 2012), lo que ha dado pie a la imagen del *ciudadano del mundo*; es decir, el comportamiento del individuo ha dejado de estar determinado por el sentimiento de pertenencia a un lugar específico, de manera que los hablantes de lenguas minoritarias reclaman el prestigio de sus variedades maternas e incluso el derecho de sus hijos a hablarlas.

Esta nueva situación sociopolítica se refleja, como no podría ser de otra manera, en el comportamiento lingüístico. Considera Patiño-Santos que el análisis etnográfico de las narrativas permite superar este escollo metodológico, puesto que hace posible negociar

identidades y relaciones de poder. La reflexividad es una forma de autoconciencia que permite al individuo actuar en relación con sus circunstancias sociales (Pérez-Milans 2017: 2). En este sentido, las narrativas permiten una doble reflexividad: cuando el hablante narra su historia y cuando el investigador repiensa continuamente sobre lo que percibe en la narración.

En segundo lugar, dado que el objetivo del trabajo es analizar las narrativas de los migrantes en relación con el contexto etnográfico, la autora detalla que el material ha sido obtenido a lo largo de cuatro años a partir de tres adolescentes de origen latinoamericano, alumnas del colegio Evangelista: un centro escolar situado en Madrid (España) que experimentó un descenso del alumnado local que concluyó con su cierre como consecuencia de un aumento constante de la densidad migratoria y su consecuente estigmatización social. Considera, pues, Patiño-Santos que la actuación comunicativa de estas alumnas está ligada a esta realidad etnográfica.

El ejemplo que presenta (*La rubia que nos llamó “idiotas”*) consiste en una situación en la que las tres alumnas fueron insultadas por una profesora y en cómo la narración compartida de esta situación creó un sentimiento de solidaridad entre ellas. Concluye, además, que la cercanía que pudo lograr la investigadora resultó una gran ventaja, no solo para la obtención de materiales, sino también para facilitar la reflexividad de las entrevistadas, la cual se materializó discursivamente en el hecho de que se consideraron con la capacidad de evaluar moralmente a su oponente (la profesora).

Del presente estudio, cabe resaltar la necesidad de reconsiderar la importancia y el papel de la postura del propio investigador frente a los actores sociales con los que interactúa durante el desarrollo del trabajo, un concepto fundamental en la sociolingüística de corte etnográfico. No se debe obviar que cada actuación y comportamiento del investigador influye directa e indirectamente en el comportamiento de los actores sociales.

Santiago Sánchez Moreano analiza ejemplos conversacionales en Cayena (Guayana Francesa) y Cali (Colombia), dos puntos geográficos estratégicos para la investigación del contacto de lenguas (tercer artículo). Por un lado, los peruanos que viven en Cayena se encuentran en contacto con hablantes de francés, hmong, thai, chino, portugués brasileño, etc. Por otro lado, los quichuas ecuatorianos que viven en Cali conviven con inmigrantes de Cauca, Huila, Nariño y Putumayo, Chocó y de la Costa Pacífica (afrocolombianos, indígenas, grupos rurales), entre otros. El artículo se asienta sobre dos propuestas: una teórica (los hablantes se posicionan socialmente a partir del empleo de elementos lingüísticos, y no de lenguas) y una metodológica (combinar las técnicas del análisis conversacional y de la interacción con la observación etnográfica y la teoría de *stance*).

En cuanto a la propuesta teórica, el autor considera que los contextos de superdiversidad y transnacionalismo, tan característicos hoy en día (y especialmente en las comunidades analizadas), junto con la exposición a un gran número de lenguas y de variedades (de lengua, estilísticas, registros, formas particulares de hablar, etc.) dan pie a que los individuos se posicionen socialmente a partir de repertorios lingüísticos cada día más complejos; esto es, la imagen de las lenguas como mecanismos delimitados resulta insuficiente para comprender el uso del lenguaje en las sociedades superdiversas modernas, por lo que conviene basarse en elementos lingüísticos (Jørgensen et al. 2011: 27).

En cuanto a la propuesta metodológica, Sánchez Moreano sostiene que, a pesar de que el análisis conversacional amplía el conocimiento sobre cómo los individuos utilizan los recursos lingüísticos, resulta necesario incorporar un enfoque etnográfico. Este hecho lo ejemplifica con una conversación entre dos hablantes en el mercado de Cayena en la que ambos emplean recursos lingüísticos del portugués, del español y del francés, y en la que se renegocian las posturas sociales cuando interviene una tercera persona; de esta manera, sostiene que el foco no está en las nacionalidades de los hablantes, sino en los repertorios escogidos a partir de su indexicalidad. Asimismo, partiendo de la noción de *stance* (Du Bois 2007) (la adopción de una postura social mediante el juicio de un elemento del contexto discursivo por parte de un individuo), el autor afirma que los hablantes se posicionan socialmente mediante elementos lingüísticos con significados asociados. Por ejemplo, Anita se vale del debilitamiento de /-s/ implosiva (característica local) para limar las fronteras sociales con los nativos de la comunidad.

Los cuatro interesantes ejemplos que se muestran en el artículo le permiten concluir, pues, que el análisis secuencial de la conversación propuesto por Auer (1995) no debe adaptarse en el nivel de los códigos, sino en el nivel de las características lingüísticas. Cabe destacar que el autor propone ideas complementarias para las pocas posibles lagunas metodológicas que pudiesen existir: métodos postestructuralistas, o análisis de escalas sociolingüísticas y cronotopos (Blommaert 2015).

Marleen Haboud propone una interesante perspectiva de metodología colaborativa (cuarto artículo). Su trabajo consiste en una recogida de datos en varias comunidades de Ecuador, un país donde el español convive con 13 lenguas indígenas que, sin embargo, están experimentando una fuerte tendencia al desplazamiento y, por tanto, están dando paso a una castellanización generalizada. Consecuentemente, presenta la autora el proyecto *Oralidad Modernidad*, cuyos principales objetivos consisten en documentar la situación de las lenguas indígenas y tratar de revitalizarlas a partir de una metodología colaborativa en la que los informantes pasan a ser hablantes y en los que se produce una continua escucha activa de la opinión y los comentarios de los miembros de la comunidad, de manera que el investigador pierde la condición de única figura activa en el estudio.

Los resultados logrados hasta ahora se recopilan en el libro *Jambij Yuracuna* (Plantas que curan), un léxico etnográfico en el que aparece la información de 107 plantas medicinales de la región, 214 fotografías, descripciones de los miembros de la comunidad, historias e incluso audios en su versión digital. La idea surgió debido a que que tanto jóvenes como mayores eran conscientes del desplazamiento lingüístico que estaban viviendo sus lenguas indígenas, como del progresivo olvido de los conocimientos ancestrales de su cultura, tales como los usos y empleos de las plantas medicinales. Esta situación propició el desarrollo de la metodología colaborativa, dado que los miembros comenzaron a sentirse implicados de manera natural al considerar que su aportación resultaba útil y obtenían resultados prácticos y reales. Así, investigadores y miembros de la comunidad se valieron de la aplicación *Lexique-pro* para diseñar el léxico etnográfico que, aparte de contener la información sobre las plantas, habla de la vida de la comunidad a través de las narraciones de la gente. Es decir, junto con las entradas y definiciones de las plantas con el nombre en quichua y castellano, se realizaron entrevistas relacionadas con su proceso de producción e historia. Finalmente, el proceso de validación de datos y edición constituyó una muestra más de metodología

colaborativa, ya que, por ejemplo, se optó por la ortografía empleada localmente por cuestiones identitarias.

Bien es cierto que puede considerarse que aún falta trabajo por hacer. La recogida de datos es demasiado reciente y no se ha producido ningún análisis *per se* que permita inducir conclusiones sobre situaciones de contacto de lenguas desde el punto de vista lingüístico. Sin embargo, su punto débil es a su vez su punto fuerte: el proyecto resulta prometedor y tiene un gran potencial. El valor etnográfico del proyecto es innegable, y la calidad del material obtenido propiciará unos resultados significativos, especialmente si se valen de las ventajas de las diferentes propuestas metodológicas que se plantean en el presente monográfico.

Jennifer Leeman, en lugar de proponer un nuevo proceso metodológico, lleva a cabo una revisión crítica de los escollos metodológicos que pueden encontrarse en las investigaciones sociolingüísticas cuyos materiales de trabajo se extraen de los censos e institutos oficiales de estadística (quinto artículo). Goza de especial interés y es un tema de actualidad, debido a la importancia que están teniendo hoy en día las diversas interpretaciones de los resultados de las encuestas que llevan a cabo los institutos oficiales de estadística

Se trata de un trabajo de carácter epistémico, con una documentación sólida y prolija. Plantea el problema de que, tradicionalmente, los datos de los censos se han empleado como fuente fiable de información, como fiel reflejo de la realidad socioeconómica de una determinada comunidad, ya que los organismos gubernamentales cuentan, por un lado, con un número de recursos mucho más elevados que los investigadores y, por otro lado, por la objetividad e imparcialidad que se les presupone.

No obstante, la autora sostiene que, como consecuencia de que el modo de categorizar y comprender el mundo es individual, el número, la forma, la ausencia y el contenido de las preguntas acerca de un determinado tema (de la lengua en el caso que nos ocupa) condiciona los resultados que se obtienen, de manera que no se consigue una representatividad real.

La redacción del texto está repleta de ejemplos y situaciones reales (especialmente a partir de los censos de Estados Unidos, Canadá, España, Italia y Escocia) que otorgan una gran solidez a la argumentación expuesta. El hecho de que, por ejemplo, en Estados Unidos lleve haciéndose mínimo una pregunta sobre cuestiones raciales desde el primer censo oficial en 1790, en tanto que la primera pregunta de corte lingüístico no llegó hasta 1890, es una buena prueba de qué era trascendental en ese momento en el país norteamericano. Este hecho adquiere mayor relevancia si se compara con la fronteriza Canadá, donde lógicamente el debate lingüístico se encuentra a la orden del día debido a su situación de contacto de lenguas, algo que se refleja de manera frecuente en sus censos.

Asimismo, las opciones de respuesta que encuentran los encuestados también son resultado de un laborioso trabajo con el fin de cumplir con un objetivo específico. Es el caso del censo del País Vasco, donde la elección de cuatro respuestas posibles, en lugar de dos, acerca de las competencias lingüísticas en dicha lengua aportan una concepción distinta de la identidad vasca.

De hecho, incluso las respuestas por las que optan los miembros de una comunidad son a menudo respondidas erróneamente de manera deliberada con la finalidad de influir en

la política lingüística de la región y conseguir que sus hijos se escolaricen en una lengua que consideraran más útil o prestigiosa.

Jennifer Leeman concluye acertadamente que, a pesar de lo expuesto anteriormente, los censos resultan de gran utilidad y presentan multitud de ventajas. Sin embargo, conviene desconfiar de ellos lo suficiente como para al menos cerciorarse de quién los produce y con qué intención para evitar un análisis sesgado. De lo contrario, pueden producirse malinterpretaciones, tales como sobrestimaciones de contextos de multilingüismo, como ocurre con los informes del Instituto Cervantes sobre el estado del español fuera de la Península, ya que a menudo se considera hispanohablante a individuos con una competencia muy limitada de la lengua al basarse solo en datos de censos en los que esa información no aparece reflejada.

Poner el foco en este problema metodológico es necesario para avanzar y mejorar, pero no conviene olvidar que, aunque un investigador contase con el mismo presupuesto que los institutos oficiales de estadística y se empeñase en reflejar una realidad de una manera absolutamente objetiva e imparcial, esto no sería posible. Epistemológicamente, se considera que el conocimiento es el punto en el que convergen la realidad/verdad y las creencias, por lo que la visión personal siempre determina el resultado obtenido.

Clara von Essen, por su parte, aprovecha la ventaja de combinar la metodología cualitativa y la cuantitativa en su investigación sobre la adaptación de los inmigrantes argentinos en Málaga (España) (sexto artículo). La autora basa su investigación sobre la variación del fonema obstruyente palatal /j/, realizado tradicionalmente como postalveolares sonoras [ʒ] y sordas [j] en Buenos Aires (BAS), y como aproximantes palatales sonoras abiertas [j] o fricativas palatales sonoras [j] en Málaga (MAL). Considera que la saliencia fonética de este proceso y, por tanto, la conciencia lingüística de la propia comunidad de habla permite reflejar el grado de acomodación de los argentinos.

La muestra se compone de 12 informantes clasificados por tres grupos de edad (18-25; 30-54; >55) y género (masculino y femenino), provenientes de Buenos Aires y con una experiencia mínima de 14 años en Málaga. Estos informantes fueron entrevistados por una investigadora local (15 minutos de entrevista semidirigida) y por la propia investigadora, también de origen argentino (30 minutos de entrevista semidirigida), con el fin de observar las implicaciones lingüísticas que pudiese conllevar. A continuación, en presencia de ambas, llevaron a cabo un test de pares ocultos y respondieron a un cuestionario sociorreticular. El análisis de la variable lingüística se llevó a cabo mediante un análisis acústico del número de cruces por cero (el parámetro acústico que mejor permite diferenciar estas fricativas) con Praat y mediante un análisis perceptivo a partir del etiquetado de la propia investigadora, con el fin de dar robustez al análisis. Estos análisis parten de la idea de que los individuos que muestren realizaciones percibidas como más aproximantes o con un menor número de cruces por cero tenderán a la acomodación lingüística con la comunidad de Málaga, y viceversa.

El principal punto positivo del trabajo de von Essen radica en la acertada combinación de metodología cualitativa y cuantitativa de manera complementaria. Por ejemplo, destaca cómo los hablantes inmigrantes adoptan rápidamente los prejuicios de las variedades del núcleo geográfico en el que viven. En el caso de los inmigrantes argentinos, los ejemplos extraídos de la observación directa y cualitativa ponen de manifiesto cómo estos han interiorizado el estereotipo negativo en cuanto a prestigio y

positivo en cuanto a solidaridad del español meridional, que es característico de las variedades septentrionales del español peninsular.

Asimismo, destaca la investigación por la combinación de diversas técnicas que se emplean de manera coherente: test de pares ocultos (*matched guise technique*, Lambert et al. 1960) con la ayuda de un actor profesional que interpreta diversas variedades lingüísticas, con objeto de conocer las actitudes de los hablantes; complementación de análisis acústico y perceptivo de la variable lingüística; diversos análisis estadísticos: correlación de Pearson, *t-test*, análisis de conglomerado bietápico con su correspondiente diagrama de Venn, etc., sin dejar olvidado la intercalación de observación directa (entrevista semidirigida) e indirecta (cuestionario sociorreticular).

Las conclusiones que pueden inferirse del estudio de von Essen son las siguientes. En primer lugar, la acomodación de los inmigrantes argentinos en Málaga se encuentra en un estadio avanzado, aunque concentrada especialmente entre los miembros más jóvenes de la comunidad. En segundo lugar, los métodos de clasificación estratificacionales resultan insuficientes para explicar la variación en las situaciones de contacto de lenguas y migración. De hecho, el modelo de conglomerados que realiza a partir de la suma de las variables macrosociales, de redes sociales y biográficas o de pequeña escala presentan un porcentaje de explicación de variación V-Cramer: 0,918, en tanto que el modelo basado solo en la edad y el género tan solo explicaban un 85% de variación. Por ejemplo, prueba que la edad de llegada a Málaga resulta una variable con más fuerza en el modelo multivariante que la variable etaria tradicional. De esta manera, propone la existencia de tres conglomerados sociales. El conglomerado 3, formado por los individuos que llegaron a Málaga con una edad comprendida entre 3 y 11 años, con una red social de malagueños, con actitudes positivas hacia la variedad vernacular y sin planes de regreso a Argentina en el futuro, es el grupo que presenta una clara acomodación (99,2%). En el extremo contrario se encuentra el conglomerado 1, compuesto por individuos que llegaron con más de 18 años, con planes de regresas a argentina, con actitudes negativas a la variedad vernacular y que solo se relacionan con argentinos (3,2%).

Víctor Fernández-Mallat plantea una investigación sobre paisaje lingüístico (PL) desde un punto de vista eminentemente cualitativo (séptimo artículo). Basándose en los trabajos de Carr (2017) y Fernández-Mallat y Leimgruber (2018), estudia de manera novedosa la relación de los individuos con las señales de su PL; es decir, no analiza la información que contienen las señales, sino cómo los miembros de la comunidad interactúan con ellas. El objetivo es incorporar los métodos del análisis de la conversación para conocer las actitudes lingüísticas, ya que el análisis de la conversación pretende la identificación y caracterización lingüística de conversaciones naturales y socialmente contextualizadas.

El estudio del autor se basa en la selección de 5 señales en español en la ciudad suiza de Basilea. Todas estas señales se caracterizan por pertenecer a establecimientos o instituciones con una larga presencia en la ciudad, de manera que todos los informantes las reconocían como elementos del PL de Basilea. Fernández-Mallat mostró las señales a cinco hispanohablantes que emigraron a Basilea y les preguntó qué evocaban en ellos la presencia del español en la ciudad mediante la técnica del tour virtual implementada en Albury (2008) y Fernández-Mallat y Leimgruber (2018). Estos informantes llevan desde dos (migrantes recientes) hasta 30 años (migrantes establecidos) en la ciudad y

que guardan una relación con el entrevistador. Cabe destacar que los hispanohablantes en Basilea constituyen un elevado porcentaje dentro de los migrantes en la ciudad (actualmente hay 5'000).

Son dos las principales conclusiones que extrae del análisis conversacional de las entrevistas realizadas. En primer lugar, cobra relevancia la cuestión espacio. La visión o consideración de la ciudad suiza como un elemento distante se relaciona con una valoración positiva de las señales en lengua española. De esta manera, buscan principalmente los migrantes recientes espacios donde su lengua materna se reterritorializa. En segundo lugar, el tiempo de estancia en Basilea, con la correspondiente mejora de las competencias en lengua alemana, juega un papel fundamental. Las experiencias vividas en el entorno de la ciudad y la percepción subjetiva del paso del tiempo matizan significativamente las actitudes.

Al igual que Marleen Haboud, la investigación presenta el problema de que la muestra de personas entrevistadas necesitaría poder ser ampliada en estudios futuros, tal y como el propio autor sugiere. Sin embargo, constituye un estadio preliminar relevante y necesario, en el que se vislumbra el potencial de implementar el análisis de la conversación no solo con el fin de lograr que los migrantes expresen sus actitudes, sino también para llevar a cabo su análisis.

En resumen, el presente volumen selecciona una lista de interesantes artículos que muestran coherencia alrededor de los nuevos enfoques metodológicos que deben incorporarse a los estudios de contacto de lenguas y migración. En este sentido, resulta básica la idea de la lengua como recursos. La realidad sociopolítica que experimentan los migrantes hoy en día induce a descartar las lenguas como entidades monolíticas. Por el contrario, debemos considerar que los migrantes tienen a su disposición un elevado número de recursos lingüísticos, asociados a variedades lingüísticas, diafásicas y diatópicas. El análisis de las actitudes y de la acomodación lingüística en estos supuestos debe contextualizarse en función de esta situación de superdiversidad y multiexposición lingüística.

Finalmente, resulta también de especial interés la idea de combinar metodologías cualitativas y cuantitativas para abordar el asunto de la migración. El siguiente paso posiblemente será eliminar el complemento dicotómico y, por tanto, disyuntivo, de dicha combinación. Parece quedar la impresión de que los métodos cuantitativos están reservados para los estudios de corte macrosociales y de que es necesaria la implementación de los métodos cualitativos para incorporar y tomar en cuenta la influencia de la variación individual en la selección de elementos lingüísticos. Sin embargo, estudios como el de von Essen ponen de manifiesto que la observación cualitativa de elementos individuales y biográficos también puede ser cuantificada e incorporada a modelos que expliquen un mayor porcentaje de realidad sociolingüística, de manera que estos elementos individuales sirven como correctores de los factores estratificacionales tradicionales; esto es, resulta muy acertado el intento de resaltar la importancia de métodos cualitativos, pero debe afrontarse desde una perspectiva de complementariedad.

Referencias bibliográficas

Albury, Nathan John. 2018. Linguistic landscape and metalinguistic talk about societal multilingualism. *International Journal of Bilingual Education and Bilingualism* 24.2: 207-223

Auer, Peter. 1995. The pragmatics in code-switching: A sequential approach. En L. Milroy y P. Muysken, eds. *One speaker, two languages: Cross-Disciplinary Perspectives on Code-Switching*. Cambridge: Cambridge University Press, pp. 115-135.

Blommaert, Jan. 2015. Chronotopes, Scales, and Complexity in the Study of Language in Society. *Annual Review of Anthropology* 44.1: 105-116.

Carr, Jhonni. 2017. *Signs of our times: Language contact and attitudes in the linguistic landscape of Southeast Los Angeles*. Tesis doctoral inédita. Los Angeles: University of California.

Codó, Eva; Patiño-Santos, Adriana y Unamuno, Virginia. 2012. Hacer sociolingüística etnográfica en un mundo cambiante: Retos y aportaciones desde la perspectiva hispana. *Spanish in Context* 9.2: 167-190.

Du Bois, John W. 2007. The stance triangle. En R. Englebretson (ed.), *Stancetaking in Discourse, subjectivity, evaluation, interaction*. Amsterdam / Philadelphia: John Benjamins Publishing Company, pp. 139-182.

Fernández-Mallat, Víctor y Leimgruber, Jakob R. E. 2018. Approaching language attitudes through discursive readings of the linguistic landscape: A Montreal case study. Póster presentado en *X-SCAPES 10th Linguistic Landscape Workshop*. Berna: Universidad de Berna.

Jørgensen, Jens Normann; Karrebæk, Martha Sif; Madsen, Lian Malai; Møller, Janus Spindler. 2011. Polylinguaging in Superdiversity. *Diversities* 2.13: 22-37.

Pérez-Milans, Miguel. 2017. Reflexivity and social change in applied linguistics. *AILA Review* 29: 1-14.

Olga Ivanova (Universidad de Salamanca). Reseña de Bañón, Antonio Miguel. 2018. *Discurso y salud. Análisis de un debate social*. Pamplona: Editorial Universidad de Navarra.

La comunicación en el ámbito de la salud y la enfermedad se ha convertido en los últimos años en uno de los temas de mayor interés de diferentes ramas de investigación, pero, de forma particular, del análisis del discurso. La necesidad de informar sobre las enfermedades, sus síntomas e impacto; la búsqueda de visibilizar los problemas que una enfermedad, sea del tipo que sea, pueda crear en el seno de una sociedad; o la urgencia de establecer pautas que permitan una comunicación fluida entre diferentes agentes involucrados en el ámbito de la salud, han puesto de manifiesto la inminencia de conocer *quién dice qué y cómo* sobre las diferentes afecciones de la salud y, *cómo el dónde, el cuándo y el cuánto* se dice sobre ellas actúan sobre la percepción pública en torno a diferentes enfermedades.

El libro de Antonio Bañón, *Discurso y salud. Análisis de un debate social* (2018), se acerca desde una perspectiva profunda y detallada –en parte, consecuencia de la larguísima trayectoria del profesor Bañón en el estudio del discurso público en el ámbito de la salud– a todas estas cuestiones y pone sobre la mesa una propuesta tan necesaria como innovadora: cómo entender la construcción de discurso sobre la salud y la enfermedad y cómo analizarlo.

La complejidad del discurso público sobre la salud, al que Bañón delimita como *debate social* en alusión al interés que el hablar sobre la salud y la enfermedad suscita invariablemente en una sociedad, no puede ser más evidente. Como tema siempre candente, siempre de interés, el de la salud y la enfermedad va más allá de una simple referencia a un estado físico concreto. Hace ya bastantes años (1969), el pionero trabajo de Claudine Herzlich, *Santé et maladie : analyse d'une représentation sociale*, hablaba de la indisociable relación del tema de la salud y la enfermedad con el conjunto de construcciones mentales, valores, concepciones y relaciones de la sociedad en la que se produce. Ahora, cincuenta años después, el libro de Bañón da cuenta de cuáles son los elementos que constituyen este conjunto: desde la concepción misma de las categorías *salud* y *enfermedad*, hasta las nuevas conceptualizaciones creadas a partir de ellas (pensemos, por ejemplo, en los procesos de *patologización* o *desmedicalización*) y las formas de representar las diferentes enfermedades en el discurso público.

En esta representación, juegan un papel nada desdeñable los múltiples agentes y las relaciones tan diversas creadas en torno al tema de la salud y la enfermedad en todos los planos del discurso. Por un lado, destacan los nuevos actores, como los medios de comunicación y los responsables políticos; las asociaciones de pacientes y los profesionales sanitarios; o los mismos pacientes, cada vez más implicados en el discurso de visibilización de sus propias enfermedades. Por otro lado, se crean nuevos espacios en las redes sociales y surgen nuevas herramientas comunicativas –como la atención telefónica–, que dan lugar a formatos discursivos hasta ahora no existentes: terapias discursivas, narrativas de la salud o testimonios desde abajo. Por último, resulta particularmente evidente el surgir de nuevos modelos discursivos que combinan la hibridación de las modalidades de habla (pensemos, por ejemplo, en la acomodación del registro endocomunicativo propio de los especialistas en medicina) y la hibridación de

los enfoques al estudio del discurso sobre la salud y la enfermedad, que implica necesariamente la medicina, la lingüística, las ciencias sociales o los estudios de la comunicación y la retórica. Así, pues, este ámbito de estudio se vuelve profundamente interdisciplinar y gira en torno a la gran pregunta que necesariamente lleva el estudio del discurso sobre la salud y la enfermedad al plano de *análisis crítico del discurso*: ¿cómo esta diversidad de actores, espacios y formatos construye el conocimiento social sobre las diferentes enfermedades, modulando nuestra percepción sobre su relativa importancia social, nuestras actitudes hacia cada una de ellas y nuestra imagen sobre la necesidad de investigar e invertir en su estudio? A propósito de esto, Urrea, Muñoz y Peña (2013) hablan de que el discurso sobre la salud y la enfermedad, en realidad, crea ideologías sociales y establece límites, que dan cabida a las diferentes condiciones de la salud y la enfermedad en cada sociedad en concreto.

A este discurso híbrido, que se construye como un debate social de niveles macro, meso y micro en todos los planos (actores, géneros, alcances, etc.), dedica Antonio Bañón una monografía sin precedentes, que pone en común las reflexiones más importantes sobre el discurso sobre la salud y la enfermedad. Abarca, en un esfuerzo de revisión bibliográfica singular y partiendo de su propia experiencia investigadora nada desdeñable en este ámbito, los conceptos que maneja el análisis del discurso sobre la salud y la enfermedad, las dimensiones espacio-temporales de este tipo discursivo, sus actores y los géneros que crean en torno al tema de la salud y la enfermedad y, de manera más patente, los rasgos de la argumentación empleada en este tipo de discurso con tan gran influencia en la opinión pública. A cada uno de estos fenómenos le dedica Antonio Bañón una detallada descripción, que sienta las bases teóricas para el estudio del discurso sobre diferentes dolencias y estados de afección.

Los intereses sociales o científicos no son los únicos en guiar la creación del discurso sobre la salud; lo son, incluso en mayor medida, los intereses ideológicos y económicos que rigen a los emisores en su actividad discursiva. De esta forma, no cabe otra definición a una *enfermedad* que la de ser un constructo social, determinado por el contexto y el entorno cronológico en el que surge: lo que es una enfermedad está determinado por el discurso que se crea en torno a ella en una comunidad *aquí y ahora*. Como consecuencia, surge la necesidad de incorporar el concepto de categorías intermedias que den cuenta de la clara influencia de las ideologías sobre la percepción social de qué es una enfermedad, qué no lo es y qué otras condiciones pueden considerarse como casos intermedios de una afección. Los neologismos como *preeenfermedad* (sinónimo de convalecencia), *diabesidad* (obesidad como factor de riesgo de la diabetes), *medicalización* (proceso de asignación de medicación para una condición que no se ha considerado enfermedad) o *despatologización* (proceso de eliminación de la condición de enfermedad a una patología) son una buena prueba de ello: evidencian que el espectro interpretativo del malestar es gradual y modulable por los actores discursivos. Algunos ejemplos de cómo los intereses económicos y políticos han modulado algunas de las condiciones de la salud vienen del TDAH (trastorno por déficit de atención e hiperactividad) y de la anorexia: mientras que el primero ha pasado por la *psicologización*, o la asignación a estados normales de desarrollo de la condición patológica, la segunda se ha *despatologizado*, pasando de la condición patológica al estado de trastorno de la conducta.

En ello, de forma patente, han influido los cambios recientes en las dimensiones espacio-temporales y agenciales del discurso sobre la salud y la enfermedad. La

consolidación del ciberespacio como uno de los medios de mayor influencia en la opinión pública, que abre puertas a participación inmediata de todo tipo, pero con frecuencia anónima e inexperta, puede aportar importantes beneficios informativos y afectivos, de lo que dan cuenta muchos trabajos recientes sobre el tema (el propio Bañón para las enfermedades raras, 2020; Mapelli para las relaciones médico-paciente, 2015; Carrasco y Márquez para la fibromialgia, 2008; o Ivanova para las demencias, 2020). No obstante, el internet, que se consulta por una de cada cuatro personas por temas relacionados con la salud (León y González 2017), puede causar al mismo tiempo estigmatización y proporcionar información incierta sobre las enfermedades.

En ello, argumenta Antonio Bañón, tiene especial importancia quién es el actor del acto discursivo y qué formato le da a su discurso sobre la salud y la enfermedad. Los *macroactores*, responsables políticos y medios de comunicación, pueden dirigir el discurso hacia sus propios marcos ideológicos y modular, con su forma de referirse a diferentes enfermedades a través de debates legislativos y su plasmado comunicativo, la opinión pública sobre la importancia y los valores epidemiológicos de una afección en concreto. La construcción de los discursos en torno a las pandemias da cuenta de cómo las referencias y los focos seleccionados pueden desviar la atención o incluso atenuar el efecto global de una enfermedad sobre un grupo social.

No así los *mesoactores* y los *microactores*, con menor visibilidad social pero cada vez más empoderados en su argumentación y reivindicación de la importancia social, incluso con respecto a las enfermedades menos frecuentes. A pesar del debate sobre su fiabilidad, los *mesoactores* –asociaciones de pacientes, colectivos profesionales, empresas farmacéuticas o investigadores en el ámbito de la salud– están cada vez más implicados en la creación y protagonización de géneros discursivos híbridos (pensemos, por ejemplo, en un congreso sobre una enfermedad concreta), que acercan la realidad vivida por las personas afectadas a grupos sociales más grandes. De ello, como analiza Antonio Bañón a partir de un corpus de textos periodísticos sobre las enfermedades poco frecuentes (EPF), es prueba la mayor presencia de artículos sobre la salud y la enfermedad en medios de comunicación regionales (resulta llamativo que el *Diario de Burgos*, *La Opinión de Murcia*, *El Norte de Castilla* y *La Gaceta de Salamanca* están entre los 10 primeros medios de prensa que con más frecuencia han informado sobre las EPF, a la par con periódicos de tirada nacional como *ABC* o *La Razón*). La visibilidad y el alcance de los *microactores*, personas enfermas y sus familiares, quedan reducidos a ámbito íntimos, casi intrascendentes, pero de gran relevancia para comprender la toma de decisiones que se hace en el ámbito sanitario. Valga como ejemplo ilustrativo la entrevista previa a la donación de órganos, cuya organización comunicativa y cuyas estrategias son clave para explicar por qué unos familiares deciden donar los órganos de su ser querido y otros, no.

Así, a partir de estos tres ejes, el discurso sobre la salud y la enfermedad se presenta como llamativamente poliédrico y multinivel, en el que el debate social constituye un hipergénero. Como tal, el debate social sobre la salud y la enfermedad se basa en los principios de la *argumentación*, que modulan la estructura y los mecanismos lingüísticos empleados. Está íntimamente vinculado al sistema de creencias que los diferentes actores (*macro*, *meso* y *micro*) poseen en torno a las afecciones y dolencias: son las experiencias y las expectativas acerca de las enfermedades en el caso de los afectados; representaciones sociales de las opiniones y actitudes en el caso de grupos socioculturales más grandes (pensemos, por ejemplo, en la asociación de fumar a una

clara consecuencia nociva); o las agrupaciones valorativas en el caso de los actores de macronivel. En todos ellos, la argumentación del discurso sobre la salud incorpora como un elemento esencial la *emotividad*, con frecuencia focalizada en el miedo, que convierte el discurso sobre la salud en una hibridación de lo objetivo (por ejemplo, los resultados de una prueba) y lo subjetivo (los sentimientos que estos resultados pueden desencadenar).

La argumentación sobre la enfermedad y la salud puede pasar por diferentes estrategias lingüísticas: la *simplificación* (representación de una enfermedad poco conocida a partir de otra más conocida), la *objetivación* (representación de un síntoma abstracto a partir de un símil real), la *sensibilización objetiva* (representación de la enfermedad tal como es) o la *reivindicación* (representación apelativa a la conciencia social sobre una enfermedad). Todas estas representaciones modulan la construcción lingüística del discurso sobre la enfermedad: desde la selección léxica (las palabras como *rogar*, *denunciar* o *reclamar* son cada vez más habituales en textos reivindicativos), hasta la construcción del mensaje como compromiso. En ello, la argumentación puede expresar diferentes tipos de identificación de los actores discursivos con las enfermedades reflejadas en el discurso. Los posibles ejes pueden ser adhesión-hostilidad, admiración-desconsideración, reconocimiento-indiferencia, o incluso disculpa. De esta manera, el discurso sobre la salud y la enfermedad se caracteriza sustancialmente por una *hibridación valorativa*: la combinación de opiniones, posiciones y actitudes de los actores y los receptores acerca de una dolencia.

La aplicabilidad de las bases teóricas del discurso sobre la salud y la enfermedad, expuesta exhaustivamente por Antonio Bañón, se refleja en cómo podemos (y debemos) analizar textos, concretos o en su conjunto, que representan el debate social en el ámbito de la salud y la enfermedad. El carácter argumentativo del discurso sobre la salud convierte los textos propios de él en géneros subjetivos, guiados por la persuasión, a pesar de su aparente estructura objetivadora. Incluso en textos tan aparentemente imparciales y rígidos como son los prospectos médicos, se puede trazar la argumentación manipulativa a favor de la toma de los medicamentos: se atenúan y se relativizan los efectos secundarios (el uso de adverbios como *ocasionalmente* o *excepcionalmente* en este contexto corrige la percepción del impacto negativo a la baja), o se desvían responsabilidades por cualquier imprevisto o efecto indeseado (el uso de la expresión *si ha tomado por error* apunta invariablemente a la responsabilidad del propio paciente).

No obstante, donde mejor se ve la carga argumentativa del discurso sobre la salud es en los textos sobre las enfermedades poco frecuentes (EPF), en los que Antonio Bañón es actualmente uno de los principales referentes científicos (véanse, a modo de orientación, Bañón (2007), Bañón y Requena (2014) o Bañón y Solves (2016)). La poca visibilidad de las EPF, con frecuencia ninguneadas por macroactores en el debate social, sumada a su naturaleza sintomatológica muy variable, que resta apoyos de parte de mesoactores (no todas las EPF están respaldadas por una asociación o entidad) y, a veces, excluida del discurso por los propios afectados (hablar de su propia enfermedad puede resultar complicado cuando aun no hay diagnóstico y no se sabe de qué se padece), convierte el discurso sobre las enfermedades poco frecuentes en un continuo avance de argumentación y contraargumentación. Como debate sectorial dentro del marco general del debate social sobre la salud, el discurso sobre las EPF –también llamadas *raras*, *huérfanas* o *minoritarias*– parte de la búsqueda de visibilizar la existencia de estas

enfermedades (argumentación) y de conseguir, a través de esta visibilización, su tratamiento como afecciones importantes y no secundarias, por poco frecuentes (contraargumentación). Resulta indudable que el alcance de la visibilización discursiva de la enfermedad tiene efectos inmediatos sobre su percepción y aceptación en una sociedad; las EPF son un buen ejemplo de ello.

Como ya es propio de él, Antonio Bañón se acerca con este libro a una problemática social de gran relevancia, pero también de gran olvido en el estudio del análisis del discurso y de la comunicación, por difícil y compleja. No obstante, cada vez más especialistas de ámbitos distintos señalan la necesidad de entender cómo la construcción del discurso sobre la salud y la enfermedad modula la percepción de diferentes receptores sobre el impacto individual y social de diferentes dolencias, y qué tiene que cambiar en esta construcción para conseguir objetivos tan sustanciales como la visibilización de las enfermedades, el drama de la soledad del paciente y de sus cuidadores, la necesidad de investigar o la urgencia de ayudar en la comunicación básica entre el médico y el paciente. En una publicación de hace casi treinta años, titulada “Discourse analysis: a new methodology for understanding the ideologies of health and illness”, Deborah Lupton (1992) defendía la urgente necesidad de incluir el análisis del discurso en los estudios de la comunicación en la salud en todos los niveles. Hace tan solo cuatro años, un grupo de investigadores de Estados Unidos (Evans-Agnew, Johnson, Liu y Boutain 2016) aun denunciaban el poco uso que se hace del análisis del discurso, en particular, de su vertiente crítica, para entender el efecto de la comunicación en el ámbito de la salud en las transformaciones sociales. En este sentido, resulta de especial relevancia esta aportación de Antonio Bañón a los fundamentos del estudio teórico y empírico del discurso sobre la salud y la enfermedad.

Referencias bibliográficas

- Bañón, Antonio. 2007. Las enfermedades raras y su representación discursiva. Propuestas para un análisis crítico. *Discurso y Sociedad* 1.2: 188-229.
- Bañón, Antonio. 2010. Comunicación destructiva y agresividad verbal en la red. Notas en torno al discurso sobre las enfermedades raras. *Discurso y Sociedad* 4.4: 649-673.
- Bañón, Antonio; Requena, Samantha. 2014. La representación de las enfermedades poco frecuentes en el discurso periodístico español. Propuestas para el análisis. *Sociocultural Pragmatics. An International Journal of Spanish Linguistics* 2.2: 1-42.
- Bañón, Antonio; Solves, Josep Antoni. 2016. El debate sobre las enfermedades poco frecuentes: una mirada a través de los medios de comunicación. *Mètode. Revista de difusió de la investigació* 88: 62-67.
- Carrasco, María del Carmen; Márquez, Manuel. 2008. El valor afectivo simbólico de los foros de fibromialgia. En O. Romani, C. Larrea y J. Fernández, eds. *Antropología de la medicina, metodologías e interdisciplinaridad: de las teorías a las prácticas académicas y profesionales*. Donostia: Ankulegi, pp. 31-49.
- Evans-Agnew, Robin A.; Johnson, Susan; Liu, Fuqin; Boutain, Doris M. 2016. Applying critical discourse analysis in health policy research: case studies in regional, organizational, and global health. *Policy, Politics, & Nursing Practice* 17.3: 136-146.

Herzlich, Claudine. 1969. *Santé et maladie: analyse d'une représentation sociale*. Paris: Mouton.

Ivanova, Olga. 2020. El discurso digital sobre la demencia: hibridación de géneros y registros en los foros de cuidadores de personas con la Enfermedad de Alzheimer. *Cuadernos AISPI* 16: 125-142.

León, Ángel Ignacio; González, José Francisco. 2017. El internet en la medicina. En J.F. González y A.I. León, eds. *El reto de ser médico*. México: Rafael Zúñiga Sustaita, pp. 161-174.

Lupton, Deborah. 1992. Discourse analysis: a new methodology for understanding the ideologies of health and illness. *Australian Journal of Public Health* 16.2: 145-150.

Mapelli, Giovanna. 2015. La comunicación (e)-médico/(e)-paciente en los foros de salud. En L. Chierichetti y G. Mapelli, eds. *Discurso médico. Reflexiones lingüísticas, históricas y lexicográficas*. Bergamo: Cerlis, pp. 131-150.

Urra, E.; Muñoz, A.; Peña, J. 2012. El análisis del discurso como perspectiva metodológica para investigadores de salud. *Enfermería Universitaria* 10.2: 50-57.

Mercè Pujol Berché (Université de Perpignan Via Domitia). Reseña de Kabatek, Johannes. 2018. *Lingüística coseriana, lingüística histórica, tradiciones discursivas*. Madrid / Frankfurt: Iberoamericana / Vervuert.

El libro recoge once artículos escritos por Johannes Kabatek sobre las tradiciones discursivas. En todos ellos, se observa explícitamente o en filigrana cómo el autor, uno de los discípulos más brillantes de E. Coseriu, desgrana, explicita y argumenta en torno a elementos cruciales de las diferentes aportaciones del lingüista de origen rumano a la lingüística en general y a la lingüística románica en particular. Ambos, Coseriu y Kabatek, políglotas, eruditos y conocedores de amplios saberes epistemológicos plantean cuestiones de gran importancia para avanzar en el conocimiento científico del lenguaje y de las lenguas, sin pretender hacer una lingüística universal, sino una “lingüística *lingüística*”, como así lo dijo Coseriu. El tema abordado en cada uno de los artículos es, desde diferentes ángulos, el de las tradiciones discursivas y su importancia en los estudios tanto de historia de la lengua, como de filología e incluso más allá de estas dos disciplinas.

Los once artículos que componen el libro son los siguientes.

1. Eugenio Coseriu, las tesis de Estrasburgo y el postulado de una lingüística *lingüística*
2. “Determinación y entorno”: 60 años después
3. Lingüística empática
4. Sobre usos y abusos de la terminología lingüística
5. Algunos apuntes acerca de la cuestión de la “hibridez” y de la “dignidad” de las lenguas iberorrománicas
6. Nuevos rumbos en la sintaxis histórica
7. ¿Es posible una lingüística histórica basada en un corpus representativo?
8. Tradiciones discursivas y cambio lingüístico
9. Las tradiciones discursivas del español medieval: historia de textos e historia de la lengua.
10. Tradición discursiva y género
11. ¿Cómo categorizar las tradiciones discursivas?

El libro está introducido por una nota de los editores y otra del propio autor. En la primera nota, Cristina Bleortu y David Paul Gerards mencionan la importancia del concepto de tradiciones discursivas, así como la de proponer una serie de contribuciones sobre las mismas, ya que merecen atención más allá de la lingüística histórica. Dichos editores explican el porqué del título de este libro: “*lingüística coseriana*”, porque se deben a Coseriu los primeros trabajos sobre las tradiciones discursivas; “*lingüística histórica*”, puesto que es el hilo conductor de los artículos que contiene el libro; y “*tradiciones discursivas*” –concepto introducido por Koch en 1987– como ampliación de la concepción sobre el lenguaje del autor rumano.

En la segunda nota, Johannes Kabatek puntualiza tres aspectos fundamentales. En primer lugar, la noción misma de tradiciones discursivas que, aunque no fuera usada por Coseriu, se hallaba ligada a su propia teoría, desarrollada después por sus discípulos. En segundo lugar, el contenido del libro responde a la importancia de dicha noción, que fue

definida terminológicamente por Peter Koch en 1987, en su tesis de habilitación sobre la enseñanza retórica de la *Ars dictaminis* y los productos textuales de la Edad Media italiana. Y, en tercer lugar, la importancia del concepto que ha dado lugar actualmente a muchos trabajos de investigación.

1. Eugenio Coseriu, las tesis de Estrasburgo y el postulado de una lingüística lingüística

El Archivo Eugenio Coseriu de la Universidad de Tübingen en Alemania recoge una gran parte de los discursos pronunciados por Coseriu (1921-2002) y una buena parte de sus publicaciones y escritos. En este artículo, Kabatek desarrolla sobre todo el punto de partida de los planteamientos de este gran romanista.

En los numerosos discursos que pronunció en sus investiduras como *Doctor honoris causa*, Coseriu enumeró los cinco principios que guiaron su labor:

- El principio de objetividad científica (“Decir las cosas tal como son”).
- El principio del humanismo o principio del “saber originario”: existe una diferencia entre las ciencias humanas y las naturales. Las hipótesis en las primeras no son arbitrarias, sino derivadas de nuestro saber originario. Accedemos al lenguaje y al mismo tiempo estamos condicionados por él.
- El principio de la tradición: el saber avanza, pero se parte de lo que se ha dicho y pensado en la historia de la humanidad.
- El principio del antidogmatismo: hay que tener una postura crítica ante cualquier doctrina preconcebida.
- El principio del bien público o de la responsabilidad social: el lingüista tiene que ser capaz de contestar las preguntas que le hace el hablante para que, con su saber, se pueda hablar mejor, traducir, interpretar e incluso poder criticar, allí donde se manipula.

Coseriu consideraba que su aportación más importante para la lingüística fue la tripartición entre sistema, norma y habla o las dimensiones de las variedades, así como la diferencia entre tres planos: el universal, el histórico y el individual, cada uno de ellos considerados desde tres puntos de vista diferentes: actividad, saber y producto.

El otro aspecto fundamental es su concepción sobre el lenguaje marcada por la influencia del idealismo alemán (Herder 1772 y Humboldt 1836), de la filosofía antigua y la de los siglos XIX y XX. Las diez tesis presentadas por Coseriu son las siguientes:

- Prioridad absoluta del lenguaje. Las dos dimensiones esenciales del hombre son el lenguaje (acceso a todas las posibilidades culturales) y el trabajo (construcción de un mundo apropiado a su ser físico).
- Lenguaje y cultura. El lenguaje es una actividad creadora y cultural, en cuanto tradición cultural.
- Los universales del lenguaje. El lenguaje se caracteriza por cinco universales, tres primarios, creatividad, semántica y alteridad, y dos secundarios o derivados, historicidad y materialidad. En cuanto a la creatividad, el lenguaje crea significados, es decir, signos con significaciones. De ahí que se pueda hablar de semántica. Los signos son creados por los otros, de ahí que podamos hablar de alteridad. La historicidad resulta de la creatividad y de la alteridad. La actividad lingüística se

presenta bajo formas de sistemas tradicionales, que son propios a las comunidades históricas. La materialidad resulta de la semanticidad y de la alteridad: los significados deben estar “representados” en el mundo sensible.

- Comunicación y comunidad. Comunicación de algo, pero sobre todo con alguien. El lenguaje es un hecho social y un producto, pero sobre todo es el fundamento de toda asociación humana.
- Nombrar y decir. Las dos funciones fundamentales del lenguaje son nombrar y decir. Mientras que en la función de nombrar todo es lenguaje, la función de decir es también ciencia, actividad práctica, sentimiento, arte, poesía, etc.
- Contenido del decir. Hay que diferenciar entre designación, significado y sentido. La primera es la referencia a las “cosas”. El significado es la posibilidad objetiva de designación dada en los signos de una lengua. El sentido es la finalidad de cada “decir”, el contenido propio de un discurso como tal, la finalidad del discurso en cuanto viene expresado por la significación y la designación.
- Lenguaje y poesía. El primero es siempre obra de un sujeto dotado de alteridad. El segundo, por su parte, es obra de un sujeto que se presenta como absoluto.
- Significado y ser. El primero es siempre universal y una delimitación de una modalidad del ser.
- Significado, verdad y existencia. El primero no es ni verdadero ni falso, es anterior a la distinción misma entre verdadero y falso. Solo puede ser verdadero o falso el “decir” en cuanto proposición. Un nombre muestra una modalidad del ser, pero no dice nada acerca de ella.
- Lenguaje y “cosas”. El lenguaje no es una nomenclatura, sino que nos conduce hacia un mundo ordenado de cosas. El lenguaje técnico o terminología es lo contrario del lenguaje originario: va de la designación a la significación y nombra clases reconocidas de antemano.

A partir de estas diez tesis, Kabatek concreta algunos elementos que considera claves: la llamada “lingüística *lingüística*”, cuyo centro es la lengua particular, de manera tal que vamos de la lengua a las cosas y no viceversa. Coseriu no niega las características gestálticas de las cosas, ni sigue las pautas del relativismo enunciadas por Whorf. Los signos lingüísticos no son idénticos a las cosas, ni son los nombres dados a las cosas, sino que son conceptualizaciones de las cosas, son formas abstractas que varían de una lengua a otra. Por lo tanto, la prioridad del lenguaje no son las cosas, sino su “ser”. Las cosas sin el lenguaje no tienen existencia reconocible y comunicable. El signo lingüístico definido por Saussure en el *Cours* coincide, por lo tanto, con la concepción coseriana. Lo que caracteriza el lenguaje humano es la “organización del mundo en categorías y especies” (tesis número 5). La noción de diálogo es también fundamental por el hecho de que el lenguaje es alteridad. Por lo tanto, el lenguaje es dialógico en su esencia. Hay que subrayar que la concepción coseriana se opone a una concepción del lenguaje como “hecho social” exterior al individuo (tesis número 4). La lengua no es algo autónomo. De ahí que se considere el cambio lingüístico como un aspecto del hablar desde la actividad misma de los hablantes. Como sabemos, Coseriu distingue entre sistema, norma y habla. Las tradiciones discursivas hay que entenderlas desde la base de la tesis número 6. La explicación empieza con lo universal, sigue con lo histórico, para finalizar con lo individual. Las tradiciones discursivas añaden la segunda dimensión histórica al lado de la de las lenguas. El concepto de las tradiciones

discursivas se refiere al hecho de que “hablar es también volver a decir lo ya dicho, manteniendo la tradición o alterándola” (p. 45). Para resumir, la lengua presenta una historicidad primaria, mientras que los textos presentan una historicidad secundaria. Además, no nos servimos de las formas textuales de la misma manera en todas las lenguas.

2. “Determinación y entorno”: sesenta años después

El artículo titulado “Determinación y entorno” fue publicado en el *Romanistisches Jahrbuch* en 1955. Es uno de los textos más importantes de Coseriu, en el que desarrolla la teoría de la determinación nominal y presenta las bases de la lingüística del texto a través del concepto de los entornos. Hay que entender la lingüística del texto como esa lingüística de nivel individual que luego se llamaría pragmática (aunque con diferencias en la concepción de conceptos y en sus explicaciones).

Kabatek nos propone observar cuál es el lugar de dicho texto en la trayectoria de Coseriu y su impacto en la lingüística. Coseriu escribe este texto en la época montevideana, durante la cual publica otros trabajos de suma importancia como *Sincronía, diacronía e historia*. El texto que aquí se menciona es de los más densos y ofrece la esencia de su pensamiento, aunque no tuvo tanta difusión como otras publicaciones de Coseriu, porque “es un tratado completo condensado” (p. 52). En ese momento, se interesaba por el principio de la individualización y de la individualización histórica, y la lingüística aparece como disciplina autónoma por primera vez. Es un texto radical, en el que la lingüística del hablar se opone a la lingüística de la lengua, dándole primacía, de manera tal que hay que explicar la lengua desde el punto de vista del hablar. Las estructuras de la *langue* (en el sentido de Saussure) es un aspecto del lenguaje, pero no lo es todo. Expone además lo que llamaría una lingüística integral, en la que la tarea consiste en dar a cada problema lingüístico el lugar que le corresponde.

Describe tres niveles fundamentales: (1) el nivel del hablar, bajo el aspecto de la actividad lingüística concreta en lo universal y el discurso en lo particular, (2) el nivel histórico bajo el aspecto del saber, y (3) el nivel individual, del texto bajo el aspecto del producto. Describe también tres disciplinas, que corresponden a los tres niveles: una lingüística del hablar, una lingüística de las lenguas y una lingüística del texto. En cuanto a la relación entre el texto *Determinación y entorno*, y la pragmática, Coseriu considerará que, a lo propuesto inicialmente por Austin y posteriormente por su discípulo Searle, le falta el reconocimiento de las lenguas particulares. De ahí que no haya categorías pragmáticas en cada lengua particular y que se confunda texto, lengua y hablar, es decir, que se confundan los diferentes niveles. Así, por ejemplo, la evocación puede relacionarse con la llamada “constelación discursiva”, es decir, una constelación de entornos determinada, reconocible y categorizable. En definitiva, el texto expone la primacía del hablar, propuesta actualmente vigente y la noción de entorno que, según Kabatek, resulta menos monolítica que el saber enciclopédico” mencionado por la pragmática. En cuanto al “universo discursivo”, cuya explicación Coseriu modificó ligeramente, el autor alemán propone establecer dos continuos: uno de conocimiento de signos y el otro de conocimiento de los referentes con el fin de revisar las categorías propuestas por Coseriu.

3. Lingüística empática

Kabatek destaca en este artículo la importancia de lo que él denomina “lingüística empática”, que tiene en cuenta el hecho de que los lingüistas también son hablantes y de que la lingüística, como ciencia humana, no puede tener los mismos parámetros de investigación que las ciencias naturales. Chomsky y Coseriu se sitúan en las antípodas desde diferentes puntos de vista. El primero, monolingüe, criticado en varias ocasiones por el segundo, políglota, desea que haya una lingüística universalista basándose en una sola lengua. El segundo desea llevar a cabo una lingüística, basada en la diversidad de lenguas, enfatizando el papel de una lengua particular y llevando a cabo estudios comparativos de las lenguas. El lingüista-hablante observa continuamente lo que oye, cómo se habla; parte de ese conocimiento previo, intuitivo y precientífico no para describir “hechos objetivos”, ni teorías sobre el “mero lenguaje”, sino para reconocer fenómenos “llamativos”, formular hipótesis y llevar a cabo un estudio empírico sistemático, basado en corpus auténticos y en muestras intersubjetivamente comprobables.

Demuestra posteriormente cómo la lingüística histórica, que dispone actualmente de grandes corpus y de herramientas digitales, puede llevar a cabo investigaciones de manera rápida y eficaz, pero teniendo en cuenta que el “corpus no habla”, sino que, a partir del corpus, podemos extraer, con dedicación intensa y mucho trabajo, la información que necesitamos a partir de una pregunta bien planteada y bien tematizada. La probabilidad de que se produzca o de que se haya producido un fenómeno, probabilidad que se basa en cálculos matemáticos no nos sirve para llevar a cabo investigaciones objetivas, sino que nos sirve para poder explicar dicho fenómeno, teniendo en cuenta lo que el investigador conoce del mismo, de la esencia misma del lenguaje y de la comunicación. Subraya también que el trabajo de campo requiere tiempo y paciencia y que los informantes no son “máquinas expendedoras de datos”. A modo de conclusión, Kabatek realza la importancia de la romanística y de la filología románica que ha conservado el “espíritu del romanticismo” alemán (p. 80).

4. Sobre usos y abusos de la terminología lingüística

En este artículo, el autor lleva a cabo una serie de reflexiones epistemológicas sobre la terminología científica como “parte esencial de los objetos construidos por la ciencia” (p. 81). Presenta el estatus semiótico de los términos para pasar posteriormente a explicitar la sintaxis de la terminología (sintaxis del término), los aspectos hermenéuticos de la misma (relación entre término y designación) y la apropiación terminológica (simpatética, excluyente y vulgarizante).

En cuanto a la semiótica del término científico, argumenta que la diferencia entre los nombres propios y los nombres de referencia única es conceptual. Así, el nombre común “establece la ‘identidad’ de todos los objetivos designados por él, su pertenencia a la clase o categoría” (p. 82), mientras que el nombre propio “se refiere a un único referente entre distintos referentes designables mediante un nombre común” (p. 83). Hace referencia también a los términos técnicos, que proceden de la técnica y que son propios de los lenguajes de especialidad, son productos del trabajo humano, mientras que el término científico surge, porque se descubre un objeto y se le da un nombre. Basándose en trabajos tardíos de Coseriu, relaciona los cuatro universos en los que el discurso funciona relacionándolo con cuatro tipos de nombre: (1) el universo cotidiano

(visión subjetiva de los objetos reales) con nombres como *casa*, *Madrid*, etc.; (2) el universo científico (visión objetiva de los objetos) con nombres como *diasistema* o *triángulo*; (3) el universo de la fantasía (mundo creado) con nombres como *Alonso Quijano* y (4) el universo de la fe (mundo en el que se cree) con nombres como *diablo* o *cielo*. En cuanto a la sintaxis de la terminología, Kabatek afirma que, en el caso del español, “no existe una sintaxis particular del uso terminológico” (p. 88).

En cuanto a la apropiación del término científico, esta conlleva una modificación justificada entre el término y el objeto que él designa. Existen, según el autor, tres apropiaciones diferentes. La simpatética, que consiste en “un proceso en el que se adopta una idea junto a su contexto de creación, intentando reproducir el pensamiento original” (p. 91). Recuerda que Coseriu distingue tres niveles: (1) el universal del hablar, (2) el individual de los textos y (3) el histórico de las lenguas. Separa la tradicionalidad textual de la historicidad de la lengua como sistema estructurado, de manera tal que hay textos en los cuales habría tradición más allá de los sistemas. La apropiación excluyente, relacionada con el poder, puede ser despectiva –descalificación de un término– y prospectiva –mirada hacia el futuro–. Por último, la apropiación vulgarizante consiste en utilizar un término científico para dar más peso a la argumentación. Puede ser intencional –con larga tradición– y no intencional –a veces debido a la ignorancia–.

5. Algunos apuntes acerca de la cuestión de la “hibridez” y de la “dignidad” de las lenguas iberorrománicas

El tema de este artículo se refiere tanto a la lectura de textos científicos como a discusiones con hablantes de diferentes comunidades lingüísticas. La reflexión se basa en el hecho de si las lenguas tienen o no dignidad, o mejor dicho, en si los hablantes consideran que las lenguas son dignas de ser habladas o no. Vemos, por lo tanto, que una parte de la reflexión se centra en comunidades que tienen lenguas minorizadas, como pueden ser la gallega, que J. Kabatek conoce muy bien, y, en filigrana, la catalana, que también conoce bien. Algunos hablantes, aunque también algunos especialistas de estas lenguas minoritarias y minorizadas, opinan que no les importa cómo se hablan dichas lenguas, sino que se hablen. Se defienden por lo tanto dos posturas que son opuestas: la del “purismo” y la del “antipurismo”. Esta última ha desarrollado en la actualidad, según el autor, la perspectiva de la hibridez lingüística, en cuanto a la mezcla de lenguas e hibridación referente al proceso. Aunque hibridez tiene tradicionalmente una connotación negativa, puesto que se opone a puro, homogéneo, va adquiriendo una acepción positiva para referirse a algo creativo, incluso superior a la “pureza”. Como ejemplo de ello, el autor toma las discusiones a propósito del spanglish: algunas de las posturas no dejan de ser ideológicas o ideolingüísticas (combinación de hechos lingüísticos con hechos ideológicos). Desde la perspectiva objetiva, la discusión entre la hibridez y la pureza son discusiones de fenómenos empíricos concretos, puesto que los sistemas lingüísticos históricamente están formados por elementos procedentes de varias lenguas. Si tomamos en consideración la perspectiva de los hablantes, podemos considerar (1) que el “spanglish no es una lengua de composición histórica, sino un nombre que se da a una serie de discursos que contiene elementos de dos lenguas” (p. 119) y (2) que haya escritores y poetas que lo utilicen.

Para sintetizar, la dignidad de la lengua no es tal, “es más bien la dignidad de sus hablantes, una dignidad que lleva en sí los recuerdos de las generaciones anteriores y de

los otros presentes que la comparten. La historicidad no se opone al futuro desde la hibridez actual, sin saber si este futuro jamás se producirá y sin tener que vivirlo ellos mismos” (p. 121).

6. Nuevos rumbos en la sintaxis histórica

En este artículo, su autor desarrolla el futuro de los estudios sobre sintaxis histórica, que son cada vez más numerosos. Señala algunas pautas a partir de seis pasos: (1) interacción lingüística e innovación (cambio lingüístico), (2) categorización de los elementos que cambian, (3) sistematicidad de la lengua histórica, (4) tipología y correlación entre los diferentes cambios, (4) variación interna de la lengua histórica y (6) diferentes tradiciones discursivas. En lo referente al primer aspecto –el cambio lingüístico–, menciona que los cambios “desde abajo” conviven con los cambios “desde arriba”, existiendo entre los dos un continuo de posibilidades. Así, por ejemplo, el cultismo sintáctico tiene también su propia historia en la sintaxis del español (ejemplo del ablativo absoluto o del hipérbaton en su exageración culturalista). Añade que la innovación lingüística no es cambio, pero todo cambio presupone innovación. En cuanto a las categorías, el autor reflexiona sobre si las lenguas ofrecen categorías discretas o no. Por una parte, hay categorías claramente delimitadas, mientras que hay otras que son continuas. Así, hay continuos que ofrecen cambios graduales de forma (-s final y su omisión con diferentes pasos intermedios de aspiración) o de contenido, y otros que son frecuenciales (formas más aspiradas o de omisión, frente a formas de realización de -s). En lo referente a la sistematicidad, Kabatek toma el ejemplo del artículo, el actualizador –mejor denominación que la de especificador–, y demuestra cómo su proceso de gramaticalización y su paulatina obligatorificación se acompaña de un proceso pasivo de gramaticalización de su ausencia. Subraya el hecho de que la ausencia de un elemento en el cambio lingüístico también tiene sus funciones. El cambio de un elemento también produce cambios sistémicos, como así lo ha puesto en evidencia el estructuralismo, pero sobre todo hay que describir dichos cambios como hechos sistémicos, que corresponden al uso real del lenguaje y no como una mera reducción. De este modo, los rumbos de la sintaxis histórica tienen que interesarse también por los fenómenos negativos (hechos co-textuales).

A propósito de las correlaciones y el tipo lingüístico, la tipología histórica, que es una disciplina muy compleja, debe interesarse por conocer qué cambios están relacionados entre sí, y qué principios fundamentales rigen las grandes reestructuraciones de la lengua en distintos momentos. Señala el cambio producido por el valor no específico de los nombres escuetos, como en el conocido ejemplo de *busco camarero*, *busco un camarero* y *busco a un camarero*. Aboga por tener en cuenta las constelaciones históricas –más allá de los factores internos– que pueden acelerar o frenar ciertas evoluciones. En cuanto a la variación, propone ir más allá del estudio ya realizado de los llamados textos “semiorales” del siglo XVI para hacerlo de manera sistemática en adelante, como un hecho de “continuidad de lo anterior y emergencia de lo posterior” (p. 137). En cuanto a las tradiciones discursivas, concepto que según el autor se ha difundido muy rápidamente entre los historiadores de la lengua, destaca el hecho de que la tradicionalidad de los textos es distinta de la historicidad del sistema lingüístico y que interactúa con la gramática histórica. Distingue, de acuerdo con Coseriu, tres historicidades: la primera, como sistema de posibilidades, como técnica libre del hablar; la segunda, simultánea a la primera, es la de la tradición, la repetición de texto en la que

el lenguaje es un signo repetido, portador de lazos que indican pasados concretos; y la tercera es la historicidad de los acontecimientos únicos, que corresponde al texto como individuo. En sintaxis histórica, se suele describir la primera historicidad con ejemplos de la tercera y de ahí surgen las dudas sobre la representatividad de un texto dado para una época determinada. Las tres historicidades son simultáneas. Cada texto, como ya decía Bajtín (1989), es el reflejo de otros textos y corresponde a uno o varios sistemas lingüísticos. La variación sintáctica depende de la tradición textual. Las diferentes tradiciones discursivas se influyen entre sí, prestándose elementos, de forma que inmediatez y distancia son polos del mismo continuo.

A modo de conclusión, el autor subraya los siguientes puntos: todo cambio lingüístico tiene que ser explicable; el cambio gradual puede llevar a lo categórico, que hay que explicar; los sistemas lingüísticos funcionan con valores opositivos; los cambios no aparecen de forma aislada; la sintaxis histórica tiene que tomar en consideración el continuo de variedades entre inmediatez y distancia, y necesita corpus diferenciados, que permitan reconstruir las distintas tradiciones discursivas.

7. ¿Es posible una lingüística histórica basada en un corpus representativo?

El autor desarrolla una argumentación detallada sobre la importancia de los corpus en los estudios de historia de la lengua, disciplina que, según él, está en plena ebullición. Dichos corpus diferenciados de diversidad multidimensional permitirán la reconstrucción de grandes líneas evolutivas en la historia de la lengua española y, de manera más concreta, permitirán observar las correlaciones tipológicas. A primera vista, este artículo puede parecer destructor, pero lo que se plantea en él es una verdadera línea metodológica y teórica para las investigaciones diacrónicas. Parte del hecho de que un corpus, aunque sea recogido de manera sistemática, no puede ser representativo, porque sólo corresponde a una mínima parte de la producción textual. Los corpus deben basarse en ediciones fidedignas. De este modo, Kabatek desea llamar la atención sobre el hecho de que, en algunas versiones digitalizadas, originales o copias, hay que tomar en consideración, sobre todo cuando se trata de copias, de las cuestiones relacionadas con la forma de transcribir los errores de escritura o las grafías equivocadas. Asimismo, las ediciones deben ser fechadas y seguir un riguroso criterio filológico de fidelidad al manuscrito. También llama la atención sobre la “objetividad”, tendencia observada en lingüística bajo la influencia de las ciencias naturales y la lingüística experimental. Para poder hablar de representatividad, el autor se plantea dos preguntas. La primera se refiere al objeto mismo de la lingüística histórica y a los datos que se manejan en ella, y la segunda, a las posibilidades que nos ofrecen los corpus. Basándose en Coseriu, que niega la historicidad de los actos de habla, Kabatek considera que los actos, en cuanto actos, son universales, puesto que derivan de la semiótica de la comunicación. La universalidad de los actos no impide, sin embargo, que estos adquieran distintas formas históricas, que “pueden expresarse tradicionalmente en textos más o menos fijados o mediante formas textuales más o menos establecidas” (p. 154). La tradición es signo y puede ser tradición discursiva, pero no género discursivo. Da como ejemplo el Corpus de Referencia del Español Actual de la Real Academia Española, en el que lo que se llama “oral” son textos muy variados, transcritos con omisión de elementos fundamentales de la oralidad (cf. López Serena, 2006). La reconstrucción de la lengua histórica se hace mediante un estudio filológico (posibles alusiones históricas de cada palabra, acercamiento a la intertextualidad) y con corpus multidimensionales (estudio de

una gran cantidad de variables, reconstrucción de una gama de posibilidades entre inmediatez y distancia).

De forma sintética, el autor resume la discusión a propósito de la representatividad de los corpus aportando los elementos siguientes: no existe la representatividad cuantitativa de un corpus; puede existir cierta representatividad cualitativa; las posibilidades de una lengua varían de acuerdo con las tradiciones discursivas; estas abarcan mucho más que el género, aunque también lo incluyen; las tradiciones discursivas se pueden ordenar en el eje de la inmediatez y de la distancia; las innovaciones, lo son de las tradiciones discursivas; las innovaciones en los corpus históricos son a veces cambios del límite entre lo que se escribe y lo que no se escribe; los corpus muestran lo que existe, pero no confirman la inexistencia del elemento en cuestión.

8. Tradiciones discursivas y cambio lingüístico

El concepto de tradiciones discursivas se está empleando a menudo en la lingüística románica y actualmente también se emplea en el ámbito de los estudios históricos del español de América y del portugués de Brasil. En la lingüística alemana, el término *tradiciones discursivas* parte de la distinción coseriana de los tres niveles definidos anteriormente (cf. artículos 2 y 4). A partir de la década de los años sesenta y setenta del siglo pasado, se desarrolla una auténtica lingüística del texto, que nace como disciplina propia en la década de los setenta. A partir de entonces, se desarrollan, desde varias perspectivas, los estudios sobre la textualidad y las características de los diferentes tipos textuales. A raíz de ello, Schlieben-Lange (1987) publicó *Pragmática histórica*, libro en el que su autora relaciona la oralidad y la “escrituralidad”, sentando las bases del estudio de las tradiciones discursivas. A partir de ahí, los estudios se multiplican. Se define el concepto de tradiciones discursivas de esta forma: “partiendo de la reduplicación del nivel histórico coseriano, se postula la existencia de dos factores a nivel histórico, la lengua como sistema gramatical y léxico de una lengua, y las tradiciones discursivas” (p. 166). De ahí que el concepto de historicidad deba ser ampliado y deba diferenciarse la historicidad del ser y la historicidad de las tradiciones discursivas. Ampliarlo significa que, por ejemplo, una fórmula de saludo forma parte de una tradición, como también ciertas tradiciones en los discursos políticos. El rasgo que define las tradiciones discursivas es “la relación temporal a través de la repetición de algo” (p. 168), aunque no toda repetición es necesariamente una tradición discursiva. Así pues, “no toda repetición de un elemento lingüístico constituye una tradición discursiva, pero la formación de una tradición discursiva exige la repetición de algo” (p. 168). A ello hay que añadir la evocación. El tiempo entre dos textos tiene repetición y evocación. Por tanto, Kabatek define de esta manera el concepto de tradiciones discursivas: “la repetición de un texto o de una forma textual o de una manera particular de escribir o de hablar que adquiere valor de signo propio (es, por lo tanto, significable). Se puede formar en relación con cualquier finalidad y expresión o con cualquier elemento de contenido cuya repetición establece un lazo entre actualización y tradición, es decir, cualquier relación que puede establecer semióticamente entre dos elementos de tradición (actos de enunciación o elementos referenciales) que evocan una determinada forma textual o determinados elementos lingüísticos empleados” (p. 170). Por lo tanto, una tradición discursiva tiene valor de signo, ya que es más que un enunciado. Un texto puede ser repetido de varias formas, por lo que puede pertenecer a varias tradiciones

discursivas: existe una composicionalidad sintagmática y de “la composicionalidad paradigmática derivan diferentes posibilidades de transformación” (p. 172).

De todo ello se deduce que se pueda hablar de interferencia textual, de divergencia y de convergencia. Como propone Kabatek, cuando se estudia la historia de una lengua, lo que se está estudiando son textos que se consideran representativos de los estados de lengua en un momento determinado. Propone, asimismo, la “metodología multidimensional”, aplicada por Douglas Biber (1989, 2006) en el estudio de corpus históricos del inglés. Este autor “parte de la hipótesis de lo que permite la identificación de una tradición discursiva [Biber habla de *registros* o de *tipos de textos*] es una particular combinación de elementos en un texto” (p. 175), entre ellos, proposiciones, marcadores de aspecto y tiempo, etc. Después de un análisis factorial “se mide la evolución de los diferentes factores a lo largo de la historia y en relación con las diferentes tradiciones discursivas” (p. 175). En el proyecto de la Universidad de Tübingen, a diferencia de las propuestas de Biber, se parte de un esquema cognitivo y sintáctico, basado en la teoría de Raible (1992, 2001) para analizar las “junciones”, es decir, los esquemas sintácticos de “integración” y las relaciones semánticas (finalidad, etc.). Esto permite poner de relieve la correlación existente entre los juntadores en un texto y las tradiciones discursivas, a las cual pertenece un texto. Al analizar diferentes tipos de textos, las investigaciones llevadas a cabo por el equipo alemán han permitido poner de manifiesto, por ejemplo, que las *Fazañas* presentan frases principales sin subordinación, mientras que, en el *Libro de los fueros*, predomina la subordinación. Todo ello les permitió confirmar que las diferencias textuales observadas son indicadores de diferentes tradiciones discursivas, pero también que “los esquemas de unión que presenta un texto serán relevantes para la determinación de las tradiciones discursivas a las que el texto pertenece” (p. 181). Tras poner de relieve estos aspectos, Kabatek propone sustituir “corpus diacrónico” por “corpus diacrónico multidimensional”, es decir “un corpus de textos pertenecientes a diferentes tradiciones discursivas, con posibilidad de proyección diacrónica, pero también con posibilidad de una visión diferenciada de las distintas tradiciones discursivas coexistentes a lo largo de la historia de una lengua” (p. 182). Estos análisis cuantitativos se acompañarán del análisis filológico de detalle.

9. Las tradiciones discursivas del español medieval: historia de textos e historia de la lengua

El estudio de las tradiciones discursivas en la historia de la lengua ha hecho posible que se innove en la lingüística histórica, dejando de considerarla como una disciplina monolítica, lo que ha permitido partir de la base de que, en cada momento de la historia, estamos “ante un edificio de variedades y textos” (p. 187), por lo cual hay que adoptar una visión variacionista a lo largo de toda la historia de la lengua. Tomar en consideración las tradiciones discursivas implica también considerar conjuntamente la perspectiva interna y externa, comparando informaciones extraídas de géneros distintos. Kabatek insiste en que las tradiciones discursivas no son sinónimo de género o de texto. Ello también permite que la gramática histórica y la filología se encuentren cuando estudian los romances medievales. Eso permite además poner en evidencia, en situaciones de contacto, como en el caso del castellano y el árabe, que ciertas traducciones del árabe introducen nuevas tradiciones discursivas en el siglo XIII; o también que hay tradiciones discursivas híbridas como la prosa sapiencial. Otro enfoque

que permite el estudio de las tradiciones discursivas es la descripción de características particulares y la correlación entre diferentes “elementos *sintomáticos* para llegar a una clasificación coherente de las tradiciones discursivas” (p. 195). El tercer enfoque es la relación entre las tradiciones discursivas y la historia de la lengua. A título ilustrativo, Kabatek explica cómo “el “ideal” alfonsí consiste en la elaboración de un léxico basado en tradiciones castellanas propias, aunque incitado por textos escritos en otras lenguas” (p. 197). Se observa toda una serie de tradiciones discursivas en la lengua de Castilla. La historia de la lengua y, en particular, la de la lengua medieval, se ve enriquecida si se toman en consideración las tradiciones discursivas y no solo el sistema lingüístico.

10. Tradición discursiva y género

Foucault (1969) ya afirmó que el lenguaje es ahistórico, mientras que los textos son históricos. Coseriu distinguió entre una primera historicidad genéticamente histórica, en la que el hablante adopta la gramática y el léxico de una lengua, que es la técnica abstracta de su producción, y una segunda historicidad, que “se manifiesta en el mismo momento en que existe una relación temporal-espacial entre enunciados, sea implícita o explícita, y es dada por la intención del hablante como oyente” (p. 202). De esta forma, el término de tradiciones discursivas se sitúa “como fundamental en la teoría del lenguaje *ante* todos los fenómenos concretos, como axioma teórico respecto al cual son posteriores todas las tradiciones concretas como fenómenos empíricos más allá de sus generalizaciones y tipificaciones” (p. 202). Los géneros discursivos de Bajtín (1989) y su crítica al estructuralismo pueden considerarse, según Kabatek, como precursores de la idea de tradiciones discursivas.

Los cuatro universos del discurso (de lo cotidiano, de la ficción, de la religión y de la ciencia) propuestos por Coseriu (2006) son “mundos” culturales y construidos, son contornos en los cuales se sitúan los textos o discursos. Estos surgen de otros ya escritos, en los que coexisten distintos niveles de tradicionalidad. El “principio de la composicionalidad tradicional” es fundamental en la teoría de las tradiciones discursivas, de modo que, según este principio, “un texto puede corresponder, simultáneamente, a toda una serie de tradiciones copresentes” (p. 208), por lo que el concepto de tradiciones discursivas es mucho más amplio que el de género.

11. ¿Cómo categorizar las tradiciones discursivas?

A veces se tiene la impresión, a pesar de que el número de investigaciones sobre las tradiciones discursivas es importante, de que nos encontramos ante un cajón de sastre. Kabatek reflexiona en este artículo sobre la categorizabilidad, concretamente, sobre las categorizaciones y sobre las etiquetas de *enérgeia* discursiva, es decir, sobre la actividad creadora de textos o discursos, y sobre el *ergon* discursivo, que es el producto de dicha actividad. Subraya el hecho de que no hay que partir de categorías ya establecidas, puesto que, si se hace, se dificulta estudiar la posibilidad de que en un solo texto existan una multitud de tradiciones discursivas. Por ejemplo, en *El Conde Lucanor*, obra que la literatura categoriza como *exempla*, se encuentran evocaciones de la tradición de los fueros medievales (oraciones condicionales). Se puede llegar a la tradicionalidad estudiando, a partir de la empatía del experto, como hablante que es, y el análisis a partir de corpus, la búsqueda sistemática de lo tradicional identificando las posibles tradiciones. Lo que defiende el autor es partir desde abajo sin partir de géneros o tipos ya establecidos. Esto puede llevarse a cabo de forma horizontal o sintagmática

(diferentes párrafos evocando tradiciones distintas) o de forma horizontal o paradigmática (varios elementos que hacen referencia a varias tradiciones).

Para poder establecer categorías desde el interior del propio texto, Kabatek propone seis categorías: (1) las tres primeras, son categorías formales: (a) la forma discursiva o estructura del texto visible; (b) la zona discursiva o los diferentes apartados constitutivos, y (c) la fórmula discursiva o la repetición directa y variada del material lingüístico; y (2) las otras tres, son categorías de contenido: (a) el dominio discursivo o prácticas culturales, (b) el tema discursivo o la función central del texto, y (c) el motivo discursivo o la identificación de los topoi y las implicaturas convencionalizadas. Estas categorías pueden combinarse y darse simultáneamente constituyendo lo fundamental de la actividad del hablar.

Para terminar esta reseña, queremos señalar que estamos ante un libro que se convertirá en una obra de referencia, porque explica claramente el concepto de tradiciones discursivas y da, al mismo tiempo, pautas claras para llevar a cabo investigaciones, tanto en historia de la lengua, como en filología, a partir de la intuición del investigador como hablante-lingüista, para posteriormente verificar dicha intuición atendiendo a los instrumentos metodológicos de que se dispone como son los corpus diacrónicos multimodales.

Referencias bibliográficas

- Bajtín, Mijail, 1989. *El problema de los géneros discursivos*. Madrid: Siglo XXI.
- Biber, Douglas; Conrad, Susan; Randi, Reppen. 1989. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, Douglas; Davies, Mark; Jones, James K.; Tracy-Ventura, Nancy. 2006. Spoken and written register variation in Spanish: A multi-dimensional analysis. *Corpora* 1.1: 7-38.
- Coseriu, Eugenio. [1958]. *Sincronía, diacronía e historia*. Madrid: Gredos, 1988, 4ª ed.
- Coseriu, Eugenio. 2006, 4ª ed. *Textlinguistik. Eine Einführung*. Edición de J. Albrecht. Tübingen: Narr.
- Foucault, Michel, 1969. *L'archéologie du savoir*. Paris: Gallimard.
- Herder, Johann Gottfried. [1772]. *Abhandlung über den Ursprung der Sprache*. Stuttgart: Reclam, 1966.
- Humboldt, Wilhelm Von. 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwicklung des Menschengeschlechts*. Berlin: Dümmler.
- Koch, Peter. 1987. *Distanz im Dictamen. Zur Schriftlichkeit und Pragmatik mittelalterlicher Brief- und Redemodelle in Italien*. Habilitationsschrift, Freiburg.
- López Serena, Araceli. 2006. Las tradiciones discursivas en la historiografía lingüística y en la historia de la lengua española. En M. Fernández y A. López, eds. *Cuatrocientos años de la lengua del Quijote*. Sevilla: Universidad de Sevilla, pp. 49-54.

Raible, Wolfgang. 1992. *Junktion: eine Dimension der Sprache und ihre Realisierungsformen zwischen Aggregation und Integration*. Heidelberg: Winter.

Raible, Wolfgang, 2001. Linking clauses. En M. Haspelmath, E. König, W. Oesterreicher y W. Raible, eds. *Language Typology and Language Universals. An International Handbook*, vol. 1. Berlin/New York: De Gruyter, pp. 590-617.

Schlieben-Lange, Brigitte. 1987. *Lingüística pragmática*. Madrid: Gredos.