Enhancing university communication: an Artificial Intelligence-based tool for gender-inclusive language*

Daniela Vellutino, Mafalda Ingenito and Giuliana Vitiello (Salerno)

Abstract

The objective of this paper is to present the research carried out as part of the HCAI4GEND project, which focuses on the development of a linguistic model aimed at creating an artificial intelligence-based tool to detect, interpret and generate linguistic solutions related to sensitive language in institutional texts in Italian, English and German.

The designed tool ensures effective user engagement through a feature-rich text editor, real-time highlighting and interactive suggestion features. The continuous learning mechanism built into the artificial intelligence back-end enables the system to adapt to evolving language norms and user preferences, ensuring long-term effectiveness and cultural relevance.

To achieve these goals, the use of high-quality language resources is critical. These resources are developed following hybrid approaches that integrate contrastive typological studies with varietistic analyses, ensuring a nuanced understanding of grammatical and social usage of gender.

1 Introduction

The term *gender-inclusive language* or *gender-sensitive language* refers to a linguistic approach that acknowledges and respects individuals from typically marginalized or invisible gender groups, such as women and non-conforming genders. This approach extends to the linguistic complexities that arise from such inclusivity (cf. Taisija 2024). It is well known that this issue has been the subject of intense debate in Europe and worldwide over the past two decades (cf. Boskovic Markovic 2014).

Research has been conducted on linguistic sensitivity to gender perspectives, with particular albeit varying attention to German, English, and Italian (cf. Formato 2019; Comandini 2021; Cameron 2021). Psycholinguistic experiments have examined, for instance, the use of masculine forms in German (cf. Braun et al. 2007; Gygax et al. 2008; De Backer/De Cuypere 2012), English (cf. Ng 1990) and Italian (cf. Cacciari/Padovani 2007), as well as the role of nouns in shaping gender representation (cf. Steiger-Loerbroks/von Stockhausen 2014). These studies

* The authors share full scientific responsibility for the article as a whole. Daniela Vellutino is the author of Sections 1, 3, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, and 3.7; Mafalda Ingenito is the author of Section 2. Section 4 was co-authored.

highlight how gendered language frequently positions the masculine form as the default standard, reinforcing societal power imbalances and contributing to the perpetuation of gender inequalities.

Consequently, addressing these biases has become a crucial challenge in fostering gender equality and promoting more inclusive social environments (cf. Handoyo/Darmahusn/Sulistyaningrum 2023).

Among the institutions that have demonstrated a significant commitment to adopting more inclusive language (cf. Vellutino 2018), universities have played a particularly prominent role. The European Union, in fact, has urged its member states to implement gender-sensitive linguistic practices within their national languages. This has led to the production of multiple institutional communication guidelines and the recommendation that institutions develop their own writing guidelines as part of the implementation of Gender Equality Plans (GEPs). These plans are considered essential documents for accessing EU research funding. However, institutional texts frequently fail to adhere to the recommendations outlined in these guidelines, leaving them largely unimplemented.

Our research proposes an architecture for an open-source tool based on artificial intelligence, specifically designed to detect and analyze gender usage in institutional texts written in Italian, German, and English. The objective is to generate linguistic solutions that enable gender-sensitive institutional communication. To achieve this goal, we have identified a specific linguistic variety for training purposes: institutional language, which is the variety used by institutions for diverse pragmatic functions. The chosen domain of use is university institutional communication, and the linguistic phenomena under investigation concern the use of grammatical and social gender. This research is of interest to multiple scientific communities and, most importantly, holds practical relevance for linguistic communities seeking to promote inclusiveness in official communication.

2 Related work

In recent years, Large Language Models (LLMs) have significantly reshaped the field of Natural Language Processing (NLP), achieving remarkable performance in previously explored domains while also unlocking new avenues for research and application.

Currently, to develop tools capable of detecting and mitigating gender bias in textual data, as well as assisting users in producing gender-inclusive texts, various NLP-based rewriting systems have been proposed. These systems are designed to automatically rephrase input text to ensure gender inclusivity, leveraging advanced language modeling techniques to promote more equitable and bias-aware linguistic outputs.

Vanmassenhove/Emmery/Shterionov (2021) propose a hybrid approach to gender-neutral rewriting for the English language, combining a rule-based method with a neural model. Specifically, they utilize both manually curated synthetic data i. e.WinoBias+ and natural data from OpenSubtitles and Reddit. The approach is developed in two main stages. Initially, a rule-based rewriter is implemented using handcrafted rules, supported by an automatic error correction tool, to generate gender-neutral versions of sentences. Subsequently, the output from the rule-based rewriter is used to train a neural rewriting model. This approach overcomes some

limitations of rule-based processing by reducing the need for extensive pre-processing and decreasing the dependency on computationally expensive tools.

Sun et al. (2021), on the other hand, focus on the use of the singular pronoun *they* and propose a rewriting task aimed at producing gender-neutral English texts. In contrast to Vanmassenhove/Emmery/Shterionov (2021) approach, Sun et al.'s model does not rely on manually labelled data, yet it achieves a word error rate of less than 1%. Their approach is characterized by using lightweight and scalable techniques, to extend this methodology in the future to accommodate various identity markers across different languages.

Unlike previous approaches, Amrhein et al. (2023) present a rewriting model trained for both English and German, designed to transform gender-biased texts into more equitable versions without relying on complex linguistic rules. This makes the method suitable for languages with complex morphology, such as German. Specifically, they utilize backward augmentation techniques, which involve generating fair versions from biased texts, and round-trip augmentation, where machine translation models are used to create biased versions of fair texts through back-and-forth translations using a pivot language (e. g., English).

Focusing on the German language, Diesner-Mayer/Seidel (2022) introduce a rule-based natural language processing system capable of handling a wide range of texts and topics in German. The system takes user input; highlights words formulated as generic masculine and generates two correction suggestions from which the user can choose. The corpus used for this study is part of the open-access TIGER corpus, consisting of newspaper articles that were manually annotated to identify and verify whether nouns and pronouns represented instances of generic masculine usage.

Italian, as a language with a highly gender-marked grammatical structure, presents several challenges in adopting gender-fair language. Frenda et al. (2024) identify three primary challenges in this regard: (1) the detection of gender-marked expressions within Italian sentences, (2) the gender-fair reformulation of these expressions by rewriting gendered terms into neutral alternatives, and (3) gender-fair text generation in the context of translation, specifically ensuring the production of gender-inclusive texts when translating from English to Italian.

To address these challenges, the authors utilize three annotated datasets. The first dataset is derived from administrative documents of the University of Brescia, the second is based on a subset of the Europarl corpus, and the third consists of data from the Neo-GATE corpus, a parallel corpus designed for gender-fair English to Italian Machine Translation evaluation.

Cerabolini/Pasi/Viviani (2024) address the challenge of automating the detection of non-inclusive language in administrative documents of Italian universities through a comprehensive and innovative approach. Firstly, they develop a dictionary of keywords that, depending on their usage, may inadvertently perpetuate gender biases or reinforce stereotypes within academic documents. To complement this, they provide two correspondence lists of masculine and feminine terms to facilitate the identification and replacement of non-inclusive expressions. Additionally, they construct a dataset consisting of examples of Italian sentences extracted from academic documents that exhibit gender inclusivity issues, paired with their corresponding rewritten versions in a gender-neutral form.

3 High level tool architecture design

The tool proposed in our research project constitutes an advanced text processing system designed to support inclusive and gender-sensitive writing through an interactive plug-in interface. This system is specifically developed to assist users in text composition and revision by providing linguistic guidance based on well-defined principles. Its primary function is to analyze user-input text, identifying expressions that do not align with the prescribed guidelines and recommendations established in the linguistic instruction repository.

This system is specifically developed to assist users in text composition and revision by providing linguistic guidance based on well-defined principles. Its primary function is to analyze user-input text, identifying expressions that do not align with the prescribed guidelines and recommendations established in the linguistic instruction repository. By doing so, the tool plays a crucial role in enhancing language inclusivity and adherence to gender-equitable communication norms. The system offers appropriate solutions by leveraging two fundamental communicative strategies:

- Lexical Substitution: The tool extracts neutralized or feminized terms from a predefined thesaurus, ensuring that word choices align with inclusive language standards.
- Parallel Expressions: It suggests alternative phrasings that enhance female visibility and promote neutralization strategies, by best linguistic practices.

Additionally, through a structured repository of institutional text models, the editor provides concrete rewriting examples that adhere to inclusivity guidelines as shown in 1 and 2. This feature facilitates text improvement either automatically or semi-automatically, allowing users to refine their writing with greater ease and consistency.

Text Type	**Original text**	**Rewritten text**
Call for application	It. *Bando di concorso per funzionari*	It. *Bando di concorso per funzionari e funzionarie*
	En. *Call for applications for officers*	En. *Call for applications for officers (female/male)*
	De. *Stellenausschreibung für Beamte*	De. *Stellenausschreibung für Beamt*innen*

Table 1: Examples of gender female for the visibility of women

Text Type	**Original text**	**Rewritten text**
Form	It. *Professori di prima e seconda fascia* […]	It. *Docenti di prima e seconda fascia* […]
	En. *Full and associate professors*	En. *Full and associate professor*
	De. *Ordentliche und außerordentliche Professoren*	De. *Ordentliche und außenordentliche Professuren*

Table 2: Examples of gender-neutral language

The high-level architecture, as illustrated in 1 consists of several interconnected modules, each with specific functionalities, as explained in the subsections below.

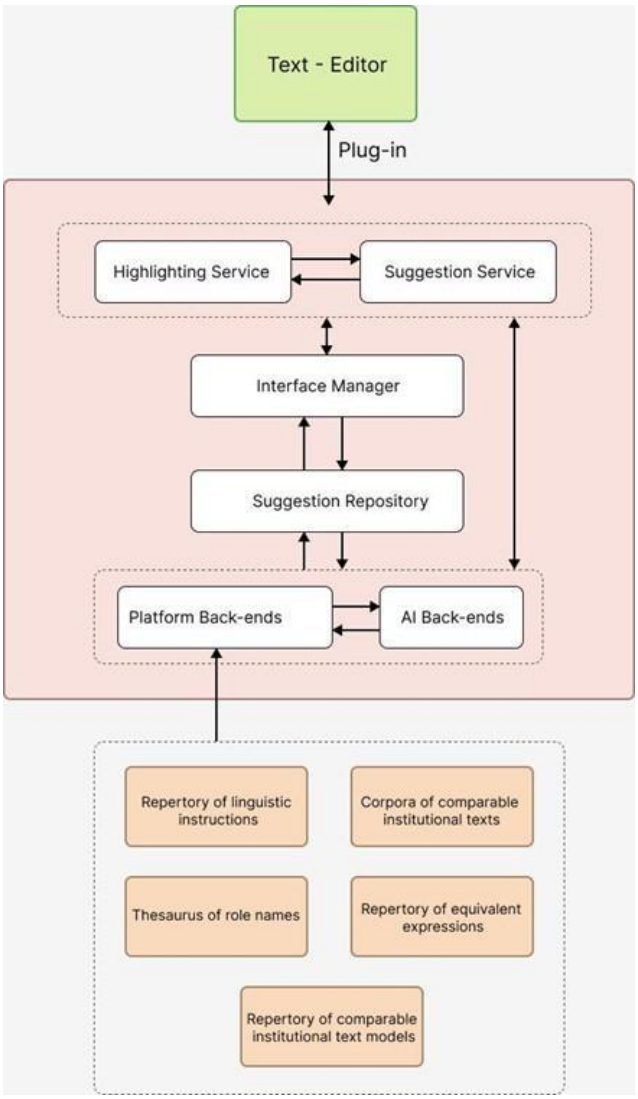

Figure 1: High-level architecture of the proposed tool

3.1 Text Editor

The Text Editor serves as the primary interface through which users interact with the tool. Rather than functioning as a simple text input field, it is a feature-rich platform specifically designed to optimize user experience, ensuring a seamless and productive workflow. Its key functionalities include:

- Real-time text composition and editing: users can type or paste text directly into the editor, allowing for immediate interaction and revision.
- Automated highlighting of gender-nonconforming terms and phrases: the system automatically detects and marks words that do not adhere to inclusive language criteria. For instance, if a user types “Gli studenti devono svolgere”, the editor will underline *studenti*, flagging it as a term requiring revision in Italian.

- Contextual suggestions: a pop-up window adjacent to the highlighted text provides alternative wordings or restructured phrases to improve inclusivity. These suggestions can be accepted with a single click, streamlining the revision process.

3.2 Highlighting Service

The Highlighting Service is responsible for identifying instances of non-gender-sensitive language by highlighting textual elements that require attention and forwarding them to the subsequent processing component. Specifically, its primary function is to scan the entire text in real time, analyze it for potential inclusivity issues, and visually flag problematic expressions for user review. The highlighting service leverages advanced NLP algorithms to perform both syntactic and semantic analysis.

This means that it goes beyond simple keyword detection, instead considering the broader linguistic and contextual framework in which terms are used, thereby identifying potentially problematic language patterns.

For example, if words such as *dottore* (‘doctor’) or *insegnante* (‘teacher’) appear in a generic context, the algorithm can recognize that, although these terms are grammatically correct, they may implicitly convey a stereotypical androcentric perspective. In such cases, the system flags them for revision, prompting the user to consider more inclusive alternatives.

3.3 Suggestion Service

The Suggestion Service provides users with tailored modifications, offering multiple options to enhance the inclusivity of their text. Once the Highlighting Service has identified non-gender-sensitive expressions, the Suggestion Service is activated.

Unlike the highlighting component, which merely flags linguistic issues, this service actively proposes specific modifications aimed at improving gender inclusivity. Rather than offering simple word-for-word replacements, the Suggestion Service employs advanced semantic and syntactic analysis to generate contextually appropriate revisions. This approach is particularly crucial in gendered language structures, where a direct substitution may not suffice to ensure inclusive and natural phrasing.

3.4 AI Back-Ends Service

The AI Back-End Service represents the core intelligence of the entire AI-driven system. It is the component where advanced processing takes place, encompassing analysis, linguistic processing, and suggestion generation through machine learning models and linguistic resources. As such, it serves as the fundamental infrastructure ensuring the system’s effectiveness in detecting and correcting non-inclusive language.

Its primary functions include:

- NLP fine-tuned models: this capability enables the system to analyze and comprehend user-written text, not only at the word level but also within the broader context of phrases and entire paragraphs.

- Machine Learning and Deep Learning Models: these models are trained on annotated datasets and linguistic corpora containing examples of both inclusive and non-inclusive language, allowing the system to learn nuanced patterns of gender-sensitive writing.

Additionally, the back end incorporates a continuous learning mechanism. Each time users interact with the system; by accepting or rejecting suggestions, the machine learning models are updated to reflect user preferences and refine the accuracy of future corrections. This adaptive learning process occurs automatically, ensuring that the system becomes increasingly effective over time as it is used.

3.5 Platform Back-Ends

The Back-End Platform serves as the foundational infrastructure that provides computational and linguistic resources powering the system's AI capabilities.

At its core lies the lingware, developed through previous and ongoing research projects focusing on institutional Italian, German, and English. This linguistic framework has been extensively analyzed about diverse pragmatic uses and communicative purposes, considering specific domains and specialized terminologies. These studies have refined an empirical research methodology, establishing a structured protocol for the development of linguistic resources following the FAIR principles (Findable, Accessible, Interoperable, and Reusable). This ensures that all resources are manually curated using standardized data and metadata formats that are machine-readable, allowing seamless integration with other software and digital services.

Specifically, these resources are:

- **Findable***:* designed with structured metadata that enables efficient data retrieval.
- **Accessible***:* optimized for open access and usability.
- **Interoperable***:* compatible with diverse computational environments, ensuring adaptability and transformation across platforms.
- **Reusable***:* exportable and importable across various software applications, facilitating integration into machine learning algorithms.

The lingware is structured into five interconnected linguistic modules, each playing a crucial role in the system's ability to process and enhance inclusive language:

a) Repository of linguistic guidelines, detailing different communicative strategies for gender-inclusive writing.
b) Parallel corpora of institutional texts, annotated both manually and semi-automatically, enriched with socio-pragmatic metadata.
c) Thesaurus of role nouns, covering institutional titles and professional designations across the three languages.
d) Repository of parallel expressions, restructured to enhance female visibility and promote linguistic neutralization.
e) Collection of modified institutional text models, aligned with strategies for gender visibility and neutralization.

This structured framework enables the AI system to offer precise, context-aware suggestions while maintaining linguistic coherence and adaptability across different institutional and professional domains.

3.6 Suggestion Repository

The Suggestion Repository is a key component of the system, responsible for storing generated suggestions, tracking user selections, and ensuring that the system delivers fast and context-aware responses. These responses are based not only on previously stored data but also on ongoing real time analyses. Additionally, this repository plays a crucial role in optimizing system performance and response speed.

In essence, whenever the system analyzes a text and generates a suggestion, the Suggestion Repository records it along with the contextual information in which it was produced. However, this storage process is dynamic rather than static: each time a user interacts with a suggestion, whether by accepting, rejecting, or modifying it, this feedback is stored in the repository. As a result, the system continuously learns from user preferences and adapts accordingly.

Over time, this iterative learning process enhances the system's efficiency, enabling it to provide increasingly accurate and user-specific recommendations, thereby improving the overall user experience.

3.7 Interface Manager

The Interface Manager is a central component of the system, acting as an intermediary and coordinator for communication between the various services and modules within the AI-driven platform. Its role is critical in ensuring seamless integration between different system components, ranging from the text editor to the AI-powered back end, while enabling the rapid and precise transfer of data.

Functioning as a centralized hub, the Interface Manager orchestrates the flow of information across the system.

Specifically, it performs the following key tasks:

- Request coordination: when a user inputs text, the Interface Manager dispatches it to the relevant services for analysis and correction. It manages incoming requests based on priority, ensuring that critical operations-such as text analysis and suggestion generation executed first.
- Data flow management: it regulates the exchange of data between the various system components, optimizing efficiency and minimizing latency.
- Bidirectional communication between front-end and back-end: it facilitates seamless interaction between the user interface and the AI-powered processing infrastructure, ensuring a smooth user experience.

By effectively managing system interactions, the Interface Manager enhances responsiveness, maintains processing efficiency, and guarantees a cohesive operation across all integrated modules.

4 Conclusion

This study has highlighted the pivotal role of gender-inclusive language in fostering equitable and respectful communication, particularly within academic institutions. By proposing the architecture of an AI-based tool to detect and mitigate gender biases in institutional texts written in Italian, German, and English, this research addresses a significant gap in existing linguistic practices.

The proposed system not only enhances gender visibility but also promotes neutralization strategies, contributing to a more inclusive and balanced representation in official communication. The advanced architecture of the tool integrates NLP techniques, such as lexical substitution and parallel expressions, to provide contextually appropriate gender-inclusive alternatives (cf. Comandini 2021).

Its design ensures effective user engagement through a feature-rich text editor, real-time highlighting, and interactive suggestion functionalities. The continuous learning mechanism embedded in the AI back end enables the system to adapt to evolving linguistic norms and user preferences, ensuring long-term effectiveness and cultural relevance.

To achieve these goals, the use of high-quality linguistic resources is crucial. These resources are developed following hybrid approaches that integrate contrastive typological studies with varietistic analyses, ensuring a nuanced understanding of grammatical and social gender usage. In subsequent phases of the project, these resources will be tested through psycholinguistic experiments using eye-tracking systems, which will contribute to validating the text editor plug-in before developing the web-based platform. This interdisciplinary approach not only enhances the accuracy and effectiveness of the tool but also ensures its adaptability across different cultural and linguistic contexts. As gender-inclusive language continues to evolve, this research lays the groundwork for future advancements in gender-sensitive NLP systems, supporting more inclusive and equitable communication practices.

References

Amrhein, Chantal et al. (2023): “Exploiting Biased Models to De-bias Text: A Gender-Fair Rewriting Model”. In: Rogers, Anna/Boyd-Graber, Jordan/Okazaki, Naoaki (eds.): *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*. Volume 1: *Long Papers*. Toronto, Association for Computational Linguistics: 4486–4506. doi: 10.18653/v1/2023.acl-long.246.

Boskovic Markovic, Valentina (2014): “The Usage of Gender Sensitive Language among Male and Female Students in Serbia”. *Anali Filološkog fakulteta* XXVI: 387–407.

Braun, Friederike et al. (2007): „Aus Gründen der Verständlichkeit...: Der Einfluss generisch maskuliner und alternativer Personenbezeichnungen auf die kognitive Verarbeitung von Texten“. *Psychologische Rundschau* 58/3: 183–189. doi: 10.1026/0033-3042.58.3.183.

Cacciari, Cristina/Padovani, Roberto (2007): “Further Evidence of Gender Stereotype Priming in Language: Semantic Facilitation and Inhibition in Italian Role Nouns”. *Applied Psycholin-guistic*s 28/2: 277–293.

Cameron, Deborah (2021): "The impact of language and gender studies: public engagement and wider communication". In: Hall, Kira/Litosseliti, Lia (eds.): *The Routledge Handbook of Language, Gender, and Sexuality*. London/New York: Routledge: 382–396.

Cerabolini, Aurora/Pasi, Gabriella/Viviani, Marco (2024): "Automating Gender-Inclusive Language Modification in Italian University Administrative Documents". In: Rapp, Amon et al. (eds.): *Natural Language Processing and Information Systems*. Cham, Springer Nature Switzerland: 333–347.

Comandini, Gloria (2021): "Salve a tuttə, tutt*, tuttu, tuttx e tutt@: l'uso delle strategie di neutralizzazione di genere nella comunità queer online. Indagine su un corpus di italiano scritto informale sul web". *Testo e Senso* 23: 43–64. testoesen-so.it/index.php/testoesenso/article/view/524/489 [24.03.2026].

De Backer, Maarten/De Cuypere, Ludovic (2012): "The interpretation of masculine personal nouns in German and Dutch: A comparative experimental study". *Language Sciences* 34/3: 253–268.

Europarl corpus. statmt.org/europarl [25.03.2026].

Diesner-Mayer, Theodor/Seidel, Niels (2022): "Supporting Gender-Neutral Writing in German". In: Mühlhäuser, Max et al. (eds.): *Proceedings of Mensch und Computer 2022. Darmstadt, Association for Computing Machinery*: 509–512. doi: 10.1145/3543758.3547566.

Formato, Federica (2019): *Gender, Discourse and Ideology in Italian*. Cham: Palgrave Macmillan.

Frenda, Simona et al. (2024): "GFG – Gender-Fair Generation: A CALAMITA Challenge". In: Dell'Orletta, Felice et al. (eds.): *Proceedings of the Tenth Italian Conference on Computational Linguistics* (CLiC-it 2024). Pisa, Italy, CEUR Workshop Proceedings: 1106–1115. https://aclanthology.org/2024.clicit-1.122/ [26.03.2026].

Gygax, Pascal et al. (2008): "Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men". *Language and Cognitive Processes* 23/3: 464–485. doi: 10.1080/01690960701702035.

Handoyo, Dharmahutama CK/Darmahusn/Sulistyaningrum, Siti Drivoka (2023): "Uncovering Linguistic Bias: A Critical Analysis of Gendered Lan-guage in the English Language". *Proceeding Internasional Conference on Child Education* 1/2: 221–235. proceeding.unimar.ac.id/index.php/icce/article/view/84 [24.03.2026].

HCAI4GEND. linfalab.cs.unisa.it/index.php/hci4gend/ [26.03.2026].

OpenSubtitles. opensubtitles.org/pb/search/subs [25.03.2026].

Neo-GATE corpus. mt.fbk.eu/neo-gate/ [25.03.2026].

Ng, Sik Hung (1990): "Androcentric coding of man and his in memory by language users". *Journal of Experimental Social Psychology* 26/5: 455–464.

Reddit. reddit.com/ [25.03.2026].

Steiger-Loerbroks, Vera/von Stockhausen, Lisa (2014): "Mental representations of gender-fair nouns in German legal language: An eye-movement and questionnaire-based study". *Linguistische Berichte* 2014/237: 56–79.

Sun, Tony et al. (2021): "They, them, theirs: Rewriting with gender-neutral English". *arXiv preprint*. doi.org/10.48550/arXiv.2102.06788 [26.03.2026].

TIGER corpus. ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger/ [25.03.2026].

Vanmassenhove, Eva/Emmery, Chris/Shterionov, Dimitar (2021): "NeuTral Rewriter: A Rule-Based and Neural Approach to Automatic Rewriting into Gender Neutral Alternatives". In: Moens, Marie-Francine et al. (eds.): *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Punta Cana, Association for Computational Linguistics: 8940–8948. doi: 10.18653/v1/2021.emnlp-main.704.

Vellutino, Daniela (2018): *L'italiano istituzionale per la comunicazione pubblica*. Bologna: Il Mulino.

WinoBias+. uclanlp.github.io/corefBias/overview [25.03.2026].