

## Semantic frames and semantic networks in the *Health Science Corpus*

Isabel Verdaguer  
Universidad de Barcelona  
i.verdaguer@ub.edu

### Abstract

The aim of this paper is to apply frame semantics principles to the analysis of a specialized corpus, the *Health Science Corpus*, implemented in the lexical database *SciE-Lex*. Taking *FrameNet* as the basis for this research, I will assign frame semantic features to *SciE-Lex* data in order to highlight the shared semantic and syntactic background of the related words in the biomedical register, give motivation to their patterns of collocates and establish frame-based semantic networks of related lexical units.

**Key words:** Health Science Corpus, biomedical English, frame semantics, FrameNet, collocational patterning

### Resumen

El objetivo de este artículo es aplicar los principios de la semántica de marcos al análisis de un corpus especializado, el *Health Science Corpus*, implementado en la base de datos léxica *SciE-Lex*. Tomando *FrameNet* como base para esta investigación, se aplica la semántica de marcos a los datos de *SciE-Lex* para destacar los aspectos sintácticos y semánticos comunes de los términos del registro biomédico, motivar sus patrones combinatorios y establecer redes semánticas basadas en marcos.

**Palabras clave:** Health Science Corpus, inglés biomédico, semántica de marcos, FrameNet, patrones combinatorios

### 1. Introduction

Corpus linguistics has allowed the analysis of lexico-grammatical patterns in a systematic way. As Johansson (2011:17) notes: "With the computational analysis tools which are now available we can observe patterns that are beyond the capacity of ordinary human observation." Empirical corpus research has shown that particular lexical classes tend to co-occur with particular structures. Lexical items occur in a limited range of patterns, which are closely linked with their meaning, and the different senses of polysemous words can be easily distinguished by the patterns in which they typically occur (Sinclair 1991). For example, the verb *argue* can be followed by a Prepositional Phrase introduced by *about* or by a *that*-finite clause, expressing different meanings. The analysis of corpus data has also revealed that particular patterns are closely associated with semantically related words (Hunston and Francis 2000). Thus, verbs which are closely related to the 'quarrel' meaning of *argue*, such as *banter* or *bicker* are also followed by a Prepositional Phrase headed by *about*, whereas verbs such

as *suggest*, *show* or *demonstrate*, related to the adduced meaning of *argue*, can take a finite *that*-clause as their complement (Atkins, Rundell and Sato 2003). In the same way, adjectives that occur in the pattern 'it + link verb + adj + clause' (*It is interesting/likely / clear/ important / true... that*) belong to some specific semantic classes, which express modality, obviousness, importance or truth. Corpus linguistics has also shown that phraseological expressions and multi-word units, which can be placed between the poles of lexicon and syntax (Nattinger and DeCarrico 1992), are very frequent. As the present paper will demonstrate, the close interrelationship between syntax and semantics can be the basis for powerful generalizations in language.

This research takes a specialized corpus, the *Health Science Corpus (HSC)*, as a starting point. The *Health Science Corpus* is a corpus of biomedical texts compiled by the research group *GreLic* with the initial aim of analyzing the language used in biomedical research articles and building the lexical database *SciE-Lex*. *SciE-Lex* in its initial stage provides morpho-syntactic, semantic and combinatory information on the non-specialized terms commonly used in biomedical discourse. In this paper I will combine the richness of corpus data with the theory of frame semantics, as the combination of data and theory allows for an exhaustive description of linguistic phenomena. Corpus data will show how an item is actually used, whereas the theoretical framework underlying the analysis provides the background against which the data can be examined and explained.

My objective is to apply frame semantics principles to the analysis of the *Health Science Corpus* and assign *FrameNet* semantic features to *SciE-Lex* data, in order to highlight the shared semantic background of the related words in the biomedical register, give motivation to their patterns of collocates and establish frame-based semantic networks of related lexical units, which will be included in *SciE-Lex* in a later stage. Frame semantics assumes that the meaning of words is best understood by reference to semantic frames, that is to say, conceptual structures or schematizations of the speaker's world that underlie their meaning. As I will show, frame semantics allows the structured organization of lexical units in terms of frames, that is, in terms of the common semantic background underlying a group of words. Headwords will be organized into frames to enhance their regular structure. With this type of organization, it is possible to facilitate the identification and understanding of all the words that belong to the same frame and express a similar sense.

I aim at showing that the *FrameNet* model is appropriate for providing a frame-based representation of the events and situations occurring in biomedical texts, and accounting for the semantic and syntactic combinatorial properties of the frame-evoking lexical units in biomedical texts. To this end, I have carried out a frame-based analysis of a selection of verbs. Verbs have been taken first because they are crucial elements controlling the whole clause. In addition, although other word classes such as nouns or adjectives can be frame-evoking words too, verbs are the most typical (Atkins, Fillmore and Johnson 2003). However, as nominalization is a typical characteristic of scientific English, nouns will also be taken into account in future research.

I also aim at identifying the collocational patterning of the lexical units which have a similar semantic and syntactic behaviour and thus belong to the same frame. Although

*FrameNet* does not specifically deal with lexical collocations, I will also examine the patterns of collocates in order to account for the collocational patterns of verbs which share syntactic and semantic characteristics. Particular attention will be paid to the semantic, syntactic and collocational differences in polysemous verbs evoking different frames. As the different meanings of polysemous words belong to different frames, the identification of frames allows the user to clearly differentiate all the meanings of polysemous frame-evoking lexical units and of their valence patterns. Thus, they can be easily distinguished.

One practical application of this study will eventually be the enhancement of *SciE-Lex* with frame-based semantic networks of related lexical units. Therefore, the dictionary user will have, in addition to an exhaustive semantic and syntactic description of the lexical units included in the dictionary, information on the interconnections among the words that belong to the same frame, and information on the different frames evoked by polysemous items. In this way *SciE-Lex* will include an onomasiological perspective, where the user can find not only exhaustive information on the use of individual words and their combinatory possibilities, but also explicit information about the words that refer to a particular situation.

This paper is structured in the following way: In the next section I will present its background and antecedents, more specifically the *Health Science Corpus* and the origins and development of *SciE-Lex*. In section 3, I will briefly introduce the theoretical principles of frame semantics and *FrameNet*. Then, I will describe the methodological framework of the Berkeley *FrameNet* project and the methodology I will use in the present paper. Next I will present and discuss a case study. In the final section, I will present the conclusions, the pedagogic and lexicographic implications of this study and future avenues.

## **2. Background and antecedents: The *Health Science Corpus* and *SciE-Lex***

The research that I am now presenting has undergone a long development, with different stages in the process. It started with the creation of a lexical database, *SciE-Lex*, carried out by the *GreLic* research group (Verdaguer, Laso and Salazar 2013). This database of non-specialized terms used in biomedical English is intended to help Spanish-speaking scientists, mainly researchers and professionals in the area of health sciences, to write and publish their papers in English, conforming to the conventions of scientific discourse. It is a tool for encoding purposes and helps the user in text production, as it provides phonetic, morpho-syntactic, semantic and collocational information.

There are three stages in the evolution of *SciE-Lex* (available at <http://www.ub.edu/grellic/eng>). In a first stage, we have included, in addition to the equivalents in Spanish, morpho-syntactic and combinatorial information, illustrated with examples and notes. In a second stage, and in line with the new tendencies of phraseological studies based on corpus, we have added prefabricated expressions (lexical bundles) and explicit information about their variability, composition, functioning and distribution in the text. Finally, in a third stage we aim to introduce frame-based information and establish semantic networks.

## 2.1. *The Health Science Corpus* and *SciE-Lex*: First stage

Since there were no specific corpora publicly available when this research was started, the first step was the compilation of a corpus that would be representative and reflect the actual use of language in scientific texts, the *Health Science Corpus* (HSC). It has approximately four million words and consists of a collection of articles from high-impact online journals that cover the disciplines of medicine, biology, biochemistry and biomedicine.

In the compilation of the corpus the texts were fully manually edited, converted into plain text files, excluding reference lists, figures, tables, names and affiliations of the authors, and stored in different folders and subfolders, according to domain and topic. Once the corpus was compiled and annotated, we used the program *WordSmith Tools* to extract a list of words, arranged alphabetically and by frequency. Terms with a frequency of less than five occurrences per million words were excluded.

Being aware that there are already several specialized dictionaries that provide the terminological equivalent of the scientific terms and that the specific terminology in English does not present a problem for the Spanish biomedical community (Verdaguer and Laso 2006), we decided to address the general words used in scientific English, because they present more combinatorial difficulties both at the syntactic and lexical levels. We gave, therefore, special prominence to verbs, which are the most important element of the sentence, around which the other elements are organized. The resulting list was compared with the Academic Word List (Coxhead 2000) and the Academic Keyword List (Paquot 2010).

After the selection of terms, using *WordSmith Tools* we extracted the lists of concordances, collocates and clusters to carry out the linguistic analysis. In the case of collocates, *WordSmith Tools* provides a list of words that appear to the right or left of the node ordered by frequencies. As for the cluster search function, the program provides sets of words, ordered from highest to lowest frequency of occurrence. The information resulting from the morpho-syntactic, semantic and collocational analysis of the corpus was stored in a database and later included in *SciE-Lex*. In this first stage, thus, *SciE-Lex* provides the following information:

- Pronunciation of each term, in audio format to help users in oral presentations.
- Word class (C). Noun (N), Adjective (Adj), Verb (V), Adverb (Adv), Preposition (P) This is the first parameter to be taken into account, since the sense, morphological characteristics and the syntactic behaviour of words are determined by their word class.
- Morphology (M). We provide morphological information on the various nominal and verbal forms, both irregular and regular (N: singular / plural; V: base form / 3rd person singular / -ing / past / participle).
- Terminological equivalence in Spanish (E), as this database is initially aimed at Spanish speakers. In polysemous words, the different equivalents have been ordered by frequency.

- Clarification of senses (*S*). In the case of polysemy, we clarify the meanings by means of a gloss or synonymous terms.
- Cross-references to related entries (*Ver*) when words are morphologically or semantically related.
- Grammatical construction (*C*). This parameter displays the patterns of occurrence in which each sense can appear. The interaction of meaning and complementation is crucial, since in many cases the different meanings of a term are expressed through different syntactic patterns. This information is essential to form a correct sentence, especially when the entry is a verb.
- List of collocates (*L*). Here we include the list of most frequent collocates, organized by lexical field and, within each field, alphabetically arranged.
- Examples of actual use (*Ex*). The selected examples illustrate and complete the information provided in the entry. These examples have been inspired by the sentences occurring in the corpus, but they have been adapted for pedagogical purposes because they are often very long and complex.
- Explanatory notes (*N*) to highlight special usages or help users to use a term in an appropriate way.

The headword *approach* will illustrate the contents of the database. Figures 1 and 2 display part of the entry for the noun and the verb *approach*. Figure 3 shows the whole entry in *SciE-Lex*. The information provided is put in a simple way, as the target users are not necessarily familiar with linguistic terminology.

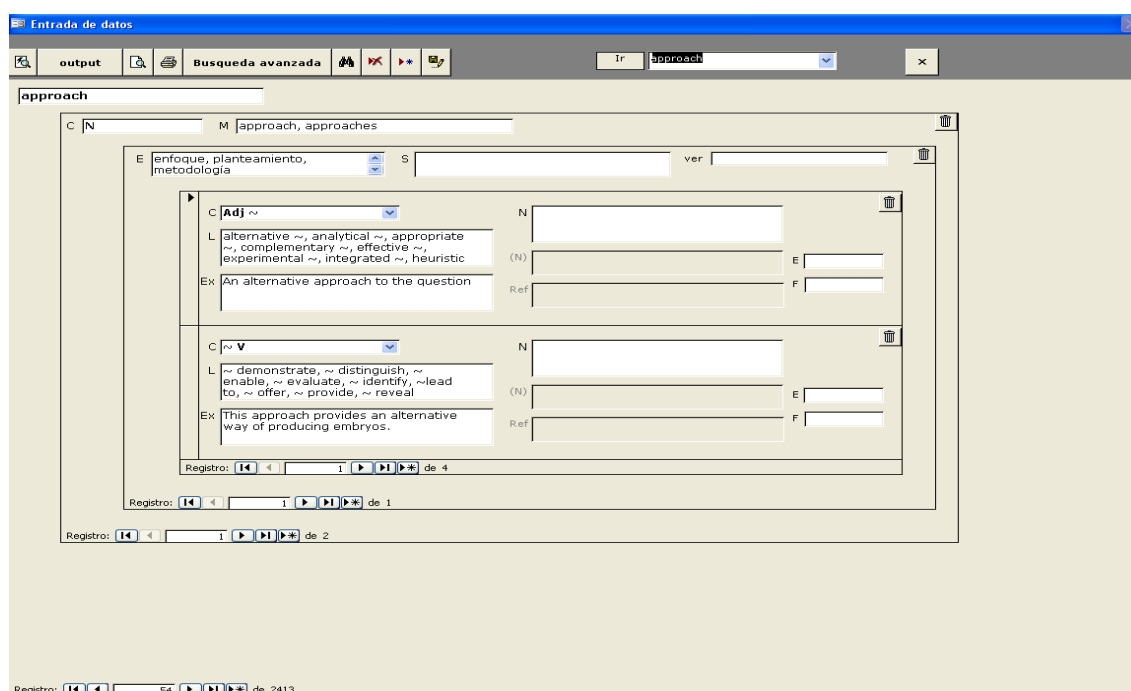


Figure 1. Database entry of the noun *approach*

As it can be seen, *SciE-Lex* provides information on the word class (*C*) of *approach* –it can be a noun (*N*)– and its morphological variants (*M*): *approach* is a countable noun and it can have a singular and a plural form (*approach, approaches*). Its Spanish equivalents (*E*) are *enfoque, planteamiento, metodología*. Next there are the patterns (*C*)

in which the noun *approach* can occur, followed by the list of the most frequent collocates (*L*) and an example (*E*) illustrating this use. Thus, *approach* can be preceded by adjectives (Adj ~): *alternative* ~, *analytical* ~, *appropriate* ~, *complementary* ~ .... It can also be the Subject of a verb (~ V): ~ *demonstrate*, ~ *distinguish*, ~ *enable* ... Examples illustrating the use of *approach* in each of its patterns, as well as clarifying notes can be added. Recursive new windows (not provided here) would show the complete complementation of *approach* as a noun: it can also occur as the object of a verb, can be followed by prepositions or by a non-finite infinitive clause.

*Approach* can also be a verb, so a new window for a different word class is generated in the database:

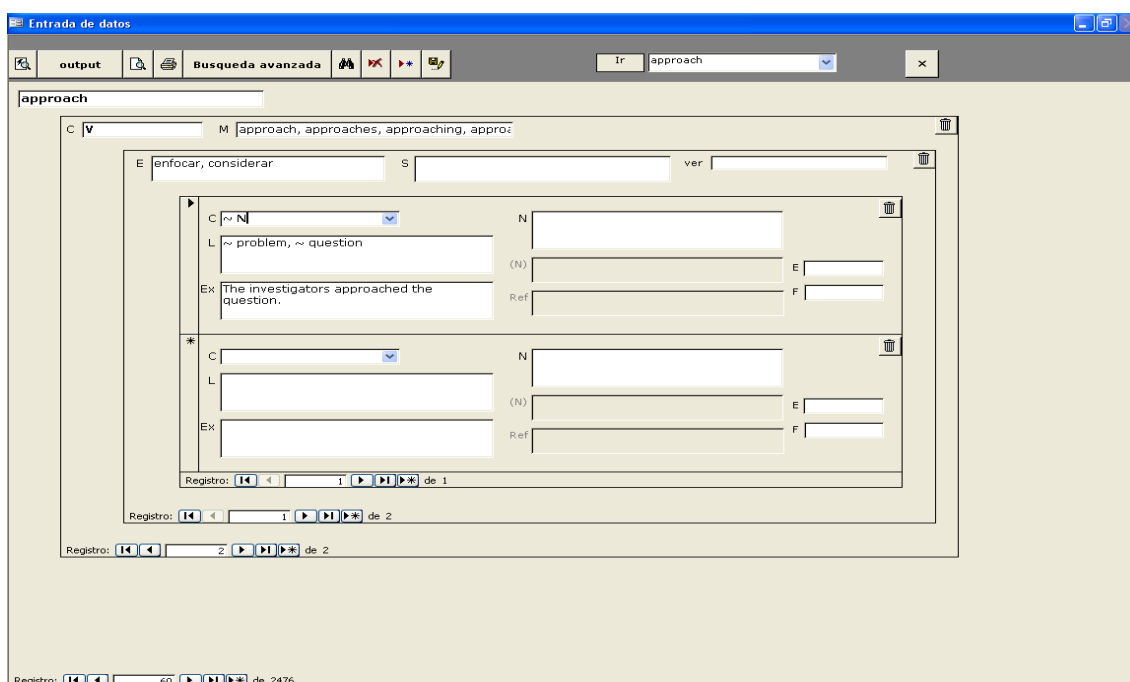


Figure 2. Database entry of the verb *approach*

As shown, the different inflected forms of the verb are displayed (*approach*, *approaches*, *approaching*, *approached*), as well as its Spanish equivalents. One of its senses is *enfocar*, *considerar*, which is transitive and can be followed by a noun as direct object (~ N). The nouns most frequently occurring in this pattern are ~ *problem*, ~ *question*. As the verb *approach* is polysemous, a new window is created in the database, providing the same type of information for this new sense. The final output of SciE-Lex is illustrated in Figure 3:

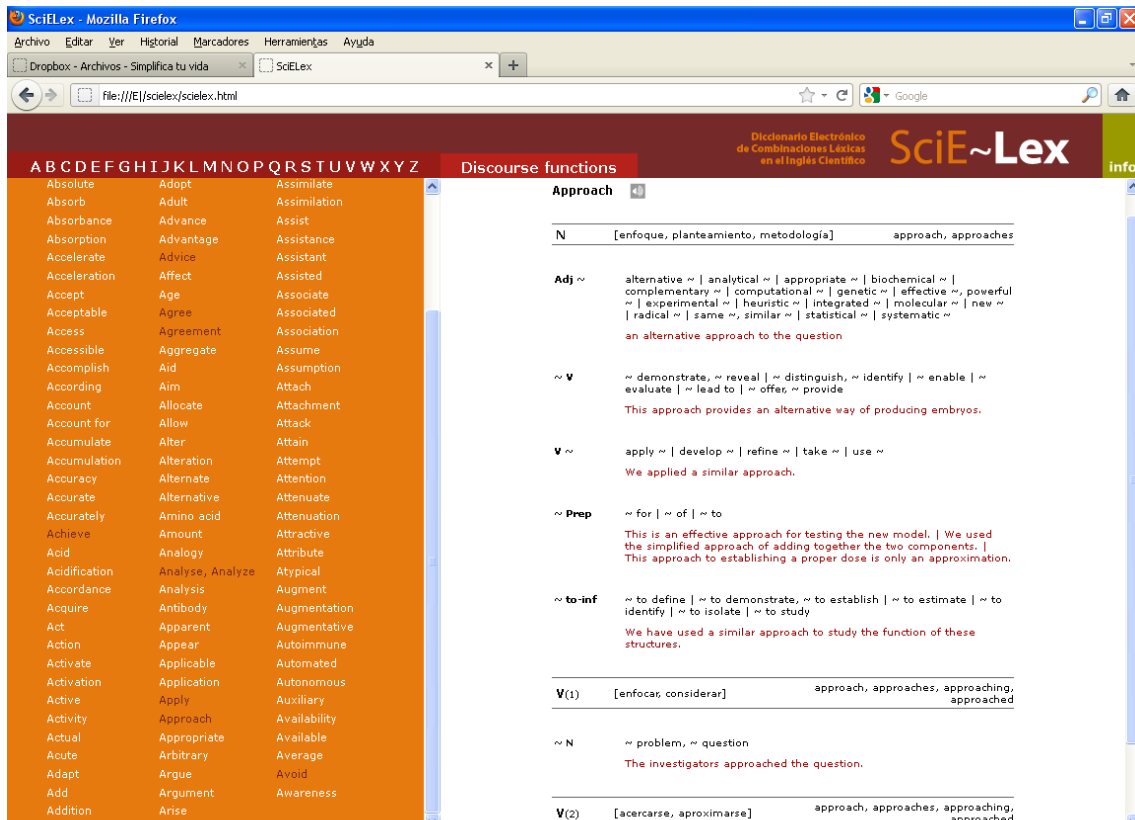


Figure 3. Final output the headword *approach* in *SciE-Lex*

## 2.2. *SciE-Lex*: Second stage

The analysis of the *Health Science Corpus* showed that phraseological patterns are not only present in everyday language but also in scientific language. As corpus linguistics research (Sinclair 1991, 2004, Stubbs 2001, Biber, Conrad and Cortes 2004, Biber and Barbieri 2007, Römer and Schulze 2009) and psycholinguistics (Nattinger and DeCarrico 1992; Wray 2002, 2008) have shown, speakers frequently use recurring combinations of words that they have stored in their brain and are important elements in the construction of discourse (Biber 2009). More recent studies (Carrió-Pastor 2017) consider that the identification of the phraseological patterns used in specific settings is a crucial issue, since phraseology is clearly register-specific (Vincent 2013). We also found that the knowledge of these units is fundamental to determine the author's membership to the scientific community, since they show their familiarity with the typical conventions of the register. As Laso and John (2013: 327) say:

NNS writers who are part of the international medical research community are committed to ensuring accurate dissemination of their research findings. This inevitably means that they need to be aware of the conventions of medical writing, so that their research articles are accepted for publication in the prestige journals of their specialized fields.

As studies carried out on learner corpora (Granger and Meunier 2008, Meunier and Granger 2008) have confirmed the difficulties involved in the use of multi-word units,

we decided to include them in *SciE-Lex*. Thus, we supplemented the lexical database with phraseological units, giving information about its composition, its discourse function and its distribution in discourse, in addition to examples of actual use and explanatory notes.

To do this, we first divided the initial corpus into several subcorpora according to the established four sections: IMRD: Introduction, Method, Results, Discussion and Conclusions. This division was carried out in order to be able to examine the composition, function and distribution of the phraseological units in each section separately.

In order to search for and select the phraseological units, we used an automatic search system with *WordSmith Tools* and *Conc-Gram*, statistical tests (mutual information) and manual revision. First, we used *WordSmithTools* to search for the sequences of three, four and five words recurring in scientific discourse. The list was later revised using *Conc-Gram*, which automatically searches for combinations of words, regardless of their position (ABD) (DBA) or the variation of their constituents (ABCD) (ABD).

At first, we decided to include in *SciE-Lex* the prefabricated structures that constitute structurally complete units of four and three words, but later on we also included other sequences that are very frequent but do not form a complete structure. Following Biber (2006), Cortes (2004) and Hyland (2008), we decided to include units of four and three words eliminating phraseological units according to two exclusion criteria: 1) three-word units that can be found in the collocational information of *SciE-Lex* and 2) we also eliminated sequences without any specific meaning or function, but frequent due to the high frequency of their individual components. A final revision of the list was made to ensure that the phraseological units included were the most useful for dictionary users.

Next, with the objective of studying the function of prefabricated structures in discourse, we established a taxonomy of discourse functions that allowed us to systematically analyze the different functions. We also analyzed what phraseological expressions present the greatest variability and what type of variability they allow, relating their variability with their discourse function and their distribution in the text.

The list of speech functions was built taking into account Biber et al.'s (2004) functional classification of lexical bundles, later modified by Hyland (2008), which classifies the phraseological units into three groups: 1) those that describe the research process (referential bundles or research-oriented); 2) those that organize discourse (discourse organizers or text-oriented); and 3) those that establish the position of the author and his interaction with the reader (stance expressions or participant-oriented). The list of discourse functions for English learners (Evans 1998) and the *MacMillan English for Advanced Learners dictionary* were also taken into account. Likewise, we included the "rhetorical moves" of Swales (1990, 2004) in our analysis.

Figures 4 and 5 show the information that we considered was necessary to include in the second phase of the dictionary:



- Phraseological unit (Lexical bundle, connected to the combinatorial dictionary entries (1st stage of *SciE-Lex*) using hyperlinks.
- Text distribution.
- Discourse function.

In addition, based on pedagogical reasons, we decided to include a field for examples extracted from the corpus and another for explanatory notes useful for the user.

The application allows you to perform different types of searches, for example, check the list of functions and see all the phraseological units associated with each function. Similarly, each phraseological unit also shows the functions it performs.

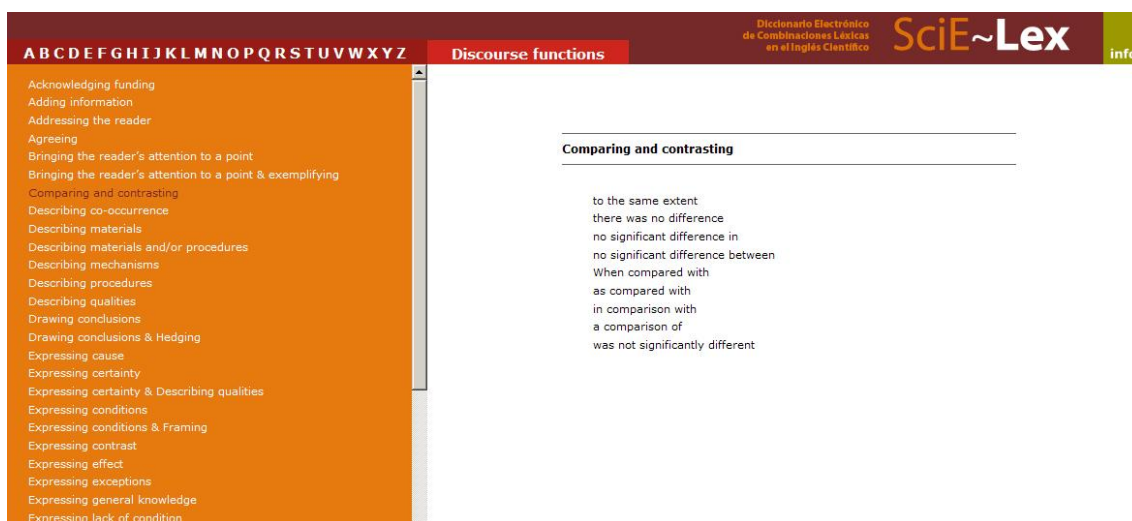


Figure 4. Discourse functions

Figure 4 shows the list of functions on the left (Acknowledging funding; Asking for information...) and the lexical bundles that have been found in the corpus expressing the function "Comparing and contrasting": *To the same extent...*

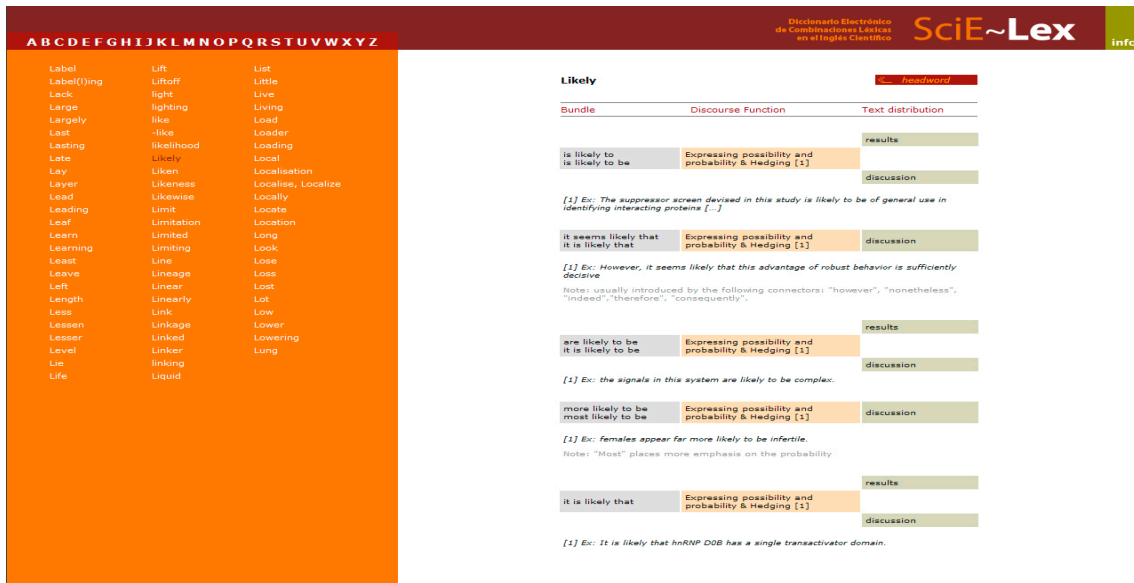


Figure 5. Lexical bundles in *SciE-Lex*

Figure 5 shows the lexical bundles containing the headword *likely*, with their discourse function and text distribution, with examples that illustrate their use.

### 2.3. *SciE-Lex*: Third stage

The third stage, which has now been started, is the application of frame semantics to the study of the biomedical register, represented by the *Health Science Corpus*. In this stage we have resorted to frame semantics and *FrameNet*, an online lexical database based on it, which systematizes the connections between related units. This new development will allow us to highlight the connections between words that have similar syntactic and semantic patterns and establish networks of frame-related lexical units. An initial search for the frames of the verbs in *SciE-Lex* has already been carried out

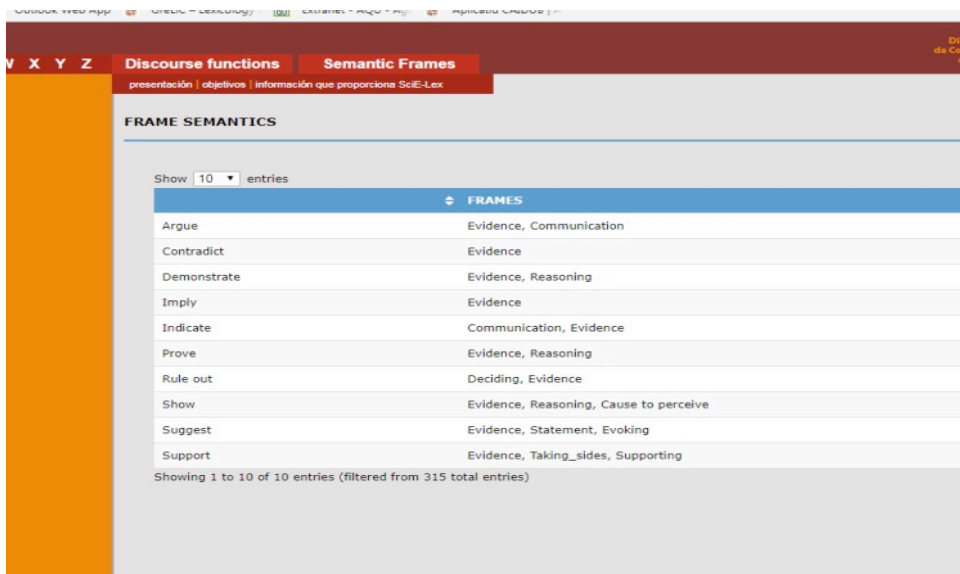


Figure 6. Semantic frames in *SciE-Lex*

### 3. Theoretical framework: Frame semantics and *FrameNet*

#### 3.1. *FrameNet*

Frame Semantics (Fillmore, 1976, 1985; Fillmore et al., 2003; Fillmore and Baker, 2010) assumes that words activate (or evoke) frames in the minds of the speakers. A frame, or semantic frame, is a conceptual structure or experience-based schematization of the speaker's world which underlies the usage of lexical units. Thus, the meaning of lexical units (LU), which, following Cruse (1986), are defined as a "pairing of a word with a sense" (Fillmore et al., 2003: 235), should be described in relation to a frame, that is to say "a schematic representation of a situation, involving various participants, props, and other conceptual roles, each of which is a frame element" (Fillmore and Petruck, 2003: 359). In these terms, a semantic frame is an essential linguistic construct for the analysis of meaning in language, since in order to understand a lexical unit the frame that it evokes, and its conceptual parts must be known.

*FrameNet* is an on-line lexical database based on frame semantics. According to Johnson and Lenci (2013: 13-14), it is "one of the major achievements in present-day research on the semantic organization of the lexicon, and on the syntax-semantics interface." The aim of *FrameNet* is to identify and define all possible frames evoked by the lexical units in a language, and analyze and annotate the sentences drawn from a linguistic corpus to show all their semantic and syntactic realizations.

*FrameNet* includes several types of linguistic information. It describes the frames underlying the different lexical units and their frame elements. It also provides lexical unit definitions and detailed information on the various syntactic realizations of semantic roles for each lexical unit, showing how this information is expressed in annotated example sentences taken from a large corpus. Information about relations between frames connecting frames to each other via semantic relations and indicating semantic relationships between connected concepts is also included.

Every sense of a word evokes a particular semantic frame or conceptual structure, which involves various participants or frame elements. For example, *argue* (see Atkins et al., 2003), as in the sentence (1), can be described with reference to the semantic frame *Quarrelling*:

(1) *They argued amicably over who should pay*

The frame *Quarrelling* involves two or more people (ARGUERS) expressing opposite ideas or beliefs about an ISSUE, the thing about which they are arguing. All the words that belong to or evoke this frame, such as, for instance the nouns *altercation*, *argument*, *disagreement* or the verbs *argue*, *bicker* or *quarrel*, are the lexical units (LU) of this frame.

In *FrameNet* the lexical units that are interpreted as having a common conceptual background belong to the same frame. A source sentence that evokes a particular frame can in principle be paraphrased by other LUs that belong to the same frame (Hasegawa

et al. 2013:114). So, synonymous expressions with different grammatical profiles can be lexical units of the same frame and it is possible to form paraphrases across different lexical categories and complementation patterns. For example, the lexical units *want* and *be eager* belong to the same frame, *Desiring*. Thus, the verb *want* is equivalent to a copula plus the adjective *eager* and the sentences *Both sides were now eager to come face to face* and *Both sides now want to come face to face* illustrate a frequent way of paraphrasing: the use of a light verb and a predicator instead of a simple verb.

Polysemous words participate in different semantic frames corresponding to their different meanings. Thus, the verb *argue*, in addition to the semantic frame *Quarrelling*, can belong to the frame *Reasoning* (2) or *Evidence* (3):

(2) *They argued that the ban was premature*

(3) *Our results argue against any systematic adverse effect of human insulin*

which have different meanings, different frame elements, and different syntactic realizations.

In the sentence *Our results argue against any systematic adverse effect of human insulin* there are not people expressing opposite ideas but results lending support to a claim *against any systematic adverse effect of human insulin*. Note too, that the noun takes different support verbs in its different meanings: *have an argument* (*Quarrelling*), *but make an argument* (*Reasoning*).

### 3.2. Frame elements

It is generally agreed that there is a need for a set of semantic roles to characterize the semantic relations of a predicate with its arguments, but there is no agreement about the number of semantic roles. Whereas Case Grammar assumes a fixed set of semantic roles, in frame semantics frame elements are described not in terms of a limited set of universal semantic roles, but in terms of the semantic frame that they evoke, thus, roles are specific for each frame. According to Lowe, Baker, and Fillmore (1997) a close examination of individual semantic fields shows the need for more detailed and fine-grained tags for semantic roles. They illustrate this condition with the following sentence:

(4) *The waters of the spa cure arthritis.*

A semantic annotation of the constituents requires at least:

- the action indicated by the verb,
- the participants (normally expressed as arguments),
- and the roles of the participants in the action.

A semantic annotation should reflect the connection between the syntactic constituents and the corresponding frame elements. In sentence (4), the grammatical subject *the waters of the spa* corresponds to the thematic causer of the cure of *arthritis*, its thematic patient and the verb's syntactic direct object. However, there is something missing in

this analysis: it does not place the event in a "generic medical event", a frame, where it would be understood that arthritis must be "borne" by some "sufferer" undergoing a treatment, which is "participating" as the patient in this event. In frame semantics terms, this event is placed in the frame *Cure*, where a HEALER treats and cures an AFFLICTION (the injuries, disease or pain) of the PATIENT.

Within each semantic frame its participants (frame elements) are identified. Some are necessary (core) although not necessarily explicit and others are peripheral or non-core, such as manner, degree, time or place. Core frame elements correspond to verb arguments in traditional grammar and non-core to adjuncts. In the example illustrating the *Quarrelling* frame the core frame elements are the ARGUERS and the ISSUE, which are the conceptually necessary participants because they are essential to the meaning of the frame. If we think about quarrelling, we have in mind someone who argues (*They*) over an issue (*over who should pay*). In addition, there may be non-core elements, such as TIME, MANNER (*amicably*), FREQUENCY or PURPOSE, which are not unique to the frame. Non-core frame elements are independent of the frame, as they are not directly related to the kind of situation described in it.

ARGUERS		MANNER	ISSUE
<i>They</i>	<i>argued</i>	<i>amicably</i>	<i>over who should pay</i>

Table I: Frame elements in the frame *Quarrelling*

The core and non-core distinction is different from the distinction between obligatory and optional elements. Whereas non-core elements are usually optional, core elements are conceptually necessary, but they may be implicit and may be left unexpressed in a given context. The grammar of a language may allow or require the omission of some element, for example the subject in imperative sentences. In the case of some verbs, like *eat*, the object may be omitted.

When frame elements are conceptually necessary but are missing from a sentence, *FrameNet* establishes three types of 'null instantiations': 'constructional', 'definite' and 'indefinite'. Constructional null instantiations are licensed by grammatical constructions; for example, omitted agents in passive constructions or the already mentioned omitted subjects of imperative sentences. For example, sentence (5):

(5) *There is no legal requirement for a child's evidence to be corroborated in civil proceedings*

illustrates the frame *Evidence*, where the SUPPORT, "a phenomenon or fact, lends support to a claim or proposed course of action, the PROPOSITION." The PROPOSITION is *a child's evidence*, while the SUPPORT (an omitted agent) is a constructional null instantiation. The omission of the agent is considered constructional because any passive sentence allows it.

Definite null instantiations can be understood in the linguistic or discourse context. In the frame *Cure* the frame element AFFLICTION is conceptually necessary since there

must necessarily be some disease which has to be cured, but it can be omitted because it can be recovered from the context, either because it had been previously mentioned or it was already known by the speakers:

(6) *The doctor cured him*

HEALER		PATIENT	AFFLICTION
<i>The doctor</i>	<i>cured</i>	<i>him</i>	CNI

Table II: Frame elements in the frame Cure

In contrast to constructional null instantiation, definite null instantiation is lexically specific. Whereas in:

(7) *We arrived at 5 pm*

the GOAL is unspecified

(8) \**We reached at 5 pm*

is not possible because the verb *reach* does not allow the omission of the GOAL.

Indefinite null instantiations are basically the implicit objects of transitive verbs such as *drink* or *eat*, which often have a specific interpretation. For instance, the missing object of *drink* is an alcoholic drink and that of *eat* is usually a meal.

### 3.3. Frame relations

*FrameNet* has built a highly structured network of frame relations, which relates frames together. Frame relations allow to connect semantically related lexical units across frames, capture generalizations and reduce the size and complexity of the lexical descriptions without losing information. The strongest relation is that of *Inheritance*, where child frames are connected to parent frames, but include additional information. The semantic facts about the parent frame also hold for the semantics of its child frames in an equally specific or more specific fact. The child frame is more specific than the parent frame but inherits all semantic properties from it. For example, the *Motion* frame, which encodes events involving a THEME “starting out in one place (SOURCE) and ending up in some other place (GOAL)” is connected to more specific frames such as *Self\_Motion*, where the theme is a living being which acts according to its own volition or *Fluidic\_Motion*, where the theme is a fluid. All the semantic roles associated with a parent frame must also be present in the child frame. For example, the SOURCE (start of the trajectory) and the GOAL (end of trajectory) are two semantic roles associated with the parent frame *Motion*. The Theme role, however, is implemented by different frame elements in the child frames: a Self mover in *Self\_Motion* and a Fluid in *Fluidic\_Motion*:

(9) *She walked along the road*

(10) *The water gushed into the house*

While in (9) the THEME is a SELF\_MOVER, a living being which voluntarily moves along a Path, in (10) the THEME is a FLUID.

Other relationships are the Metaphor or Causative\_of relations between stative frames and the corresponding Inchoative and Causative, which is frequent in the biomedical register.

### 3.4. Other *FrameNet* projects

*FrameNet* started for English. However, there are now *FrameNets* in other languages: Spanish *FrameNet* (Subirats 2009, 2013), German *FrameNet* (Burchardt et al. 2006), French *FrameNet* (Boas 2009), Japanese *FrameNet* (Sato 2009) or Swedish *FrameNet* (Borin et al. 2010), among others. Although they are all grounded in frame semantics, they may also vary from the original Berkeley *FrameNet* project in various ways. Whereas Spanish *FrameNet*, directed by Prof. Subirats, follows the Berkeley project in describing isolated corpus sentences, German *FrameNet* (SALSA) is annotating the whole corpus and French *FrameNet* uses multi-lingual corpora. They may also differ in their methodologies, particularly in the use of semi-automatic and automatic annotations.

In addition to an increasing number of *FrameNets* in languages other than English, *FrameNet* has been implemented in specific domains. As could be seen in the examples above, *FrameNet* deals with general language, but a number of studies have shown that *FrameNet* can be successfully applied to domain-specific corpora. *FrameNet*, which provides a suitable approach for the analysis of syntactic and semantic combinatorial properties of general language, can also and can provide a new perspective on specialized languages. As Dolbey (2009:93) has stated, *FrameNet* can be considered "a backbone of several domain-specific *FrameNets*."

Among the studies which have applied *FrameNet* to specialized registers, the following can be mentioned: Dolbey (2009) has applied it to molecular biology; Venturi (2013) to legal language; L'Homme (2008, 2010, 2014, 2018), L'Homme and Robichaud's (2014), L'Homme, Robichaud and Subirats (2014), L'Homme, Subirats and Robichaud (2016) to the language of ecology and computing; and Faber (2015); Reimerink and Faber (2009); Faber and Buendía Castro (2014); and Faber, León-Araúz and Reimerink (2016) to the environment and ecology. Some lexicographic devices have also applied *FrameNet* principles, among them: Schimdt's *Kicktionary* (2009), a multilingual dictionary of football terms in English, German and French; *EcoLexicon* (<http://ecolexicon.ugr.es>) a knowledge database, developed at the University of Granada, which provides *FrameNet*-like information on terms about the environment in English, Spanish and German, highlighting their relationship; or *DiCoEnviro* ([http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search\\_enviro.cgi](http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search_enviro.cgi)) and *DiCoInfo* (<http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi>), dictionaries on the language of ecology (*DiCoEnviro*) and computing (*DiCoInfo*), developed at the University of Montreal.

### 3.5 The Health Science Corpus and FrameNet

Following the current trend of applying frame semantics to specific registers and the growing body of research on FrameNet-based lexical resources for specialized language, I am taking *FrameNet* as the basis for the study of a specialized corpus of scientific English. There is a well-acknowledged need to bridge the gap between general and domain-specific language analyses, and this research represents a step towards addressing this issue. I will compare the frames and linguistic features which are present in the *Health Science Corpus* with those of the Berkeley *FrameNet* project, which is based on a corpus of general English. I will start with the general words most frequently used in biomedical texts, which are those covered by *SciE-Lex*, and in the future, I aim to continue with semitechnical and technical words.

The application of *FrameNet* principles to biomedical texts will be particularly suitable to address the main features of biomedical language, more specifically its preferred features in relation to the features of ordinary language. When frames and their frame-evoking lexical units are identified, together with their semantic and syntactic characteristics, typical and unusual meanings and valence patterns of lexical units in biomedical texts can be easily uncovered and highlighted.

The language of biomedicine is not very different from ordinary language. Apart from its specific terminology, it makes specific use of lexical and syntactic characteristics which are typical of general language. In this paper I will address the frames and the preferred meanings and syntactic patterns of the lexical units in biomedicine texts, with special emphasis on domain-specific meanings and constructions.

In specialized registers, words also usually have more specific meanings than in general language and are less polysemous. Thus, in general language the most frequent meaning of the verb *concern* and the first meaning listed in most dictionaries is "worry", as in (11):

(11) *We want to know about the issues that concern the voters*

which in FrameNet evokes the frame CAUSE\_EMOTION. However, in the *Health Science Corpus* there is no occurrence of such use. The verb *concern* always occurs evoking the frame TOPIC:

(12) *A key question concerns the cellular roles performed by each motor.*

In the present paper the characteristics of a sample of biomedical vocabulary will be identified in terms of frames and will be contrasted to those of general English. By relying on the analytical tools of frame semantics as instantiated in *FrameNet*, I will attempt to identify the realizations of events in biomedical texts and to uncover the typical and unusual meanings of lexical units in biomedical texts, and their patterns.



#### 4. Methodology

The methodology used in the different *FrameNet* projects differ slightly. Whereas the Berkeley *FrameNet* project and Spanish *FrameNet* proceed frame by frame, SALSA (German *FrameNet*) analyzes the whole corpus lemma by lemma. The methodology followed in the Berkeley *FrameNet* and in the Spanish *FrameNet* projects can be summarized in the following way:

- Identification of semantic frames and development of a frame ontology.
- Search for words that belong to the same lexical domain and bring to mind, that is to say evoke, the same frame. When a word is polysemous, it is assigned to the different frames. For example, *treat* in a medical context is associated with words such as *prevent* and is assigned to the frame *Medical\_intervention*, but in an academic context it is connected with words such as *address* and assigned to the frame *Topic*.
- Extraction from the corpus of sample sentences containing the lexical unit used with different syntactic contexts and the most frequent collocations, showing the uses of the words in the frame.
- Selection of sentences with different syntactic contexts and the most frequent collocations and annotation of the frame elements.
- Annotation of sample sentences in the corpus using the Berkeley *FrameNet* Desktop by tagging the frame-evoking lexical units with the name of the frame, and its arguments with the names of the frame elements, identifying their roles relative to the predicate and their syntactic realization. Every argument of the lexical unit is given one semantic and two syntactic labels: its frame element identity (COMMUNICATOR, TEXT...), phrase type (NP, PP...) and grammatical function (Object...). Important elements that are missing are also indicated. The relations and organization of the frames is also explicitly formulated.
- These hand tagged corpus instances are used as a training corpus to carry out the automatic tagging of examples.
- Revision of automatic annotation.

The *FrameNet* data can be accessed online at: <http://FrameNet.icsi.berkeley.edu>. *FrameNet* allows different kinds of searches: by lexical units or by semantic frames.

The methodology I have used essentially follows that of *FrameNet*. However, whereas *FrameNet* starts with frames, I have started with the lemmas of *SciE-Lex*. In addition, since my work is grounded in an existing *FrameNet*, I have analyzed the *Health Science Corpus* to detect the special features of biomedical English and uncover the specificities of this register.

- The first step has thus been an initial exploration of and identification of the frames that the verbs included in *SciE-Lex* evoke. Taking the different meanings of the verbs in *SciE-Lex* as lexical units, in an initial exploration the *GreLic* research group has manually identified the semantic frames they evoke. This has been done by checking the Berkeley *FrameNet* ontology of frames to see the different frames evoked by each lexical unit. A preliminary list of verbs and corresponding frames has been included in the *SciE-Lex* webpage, so that the

user can look up the information about the targeted lexical item in the *FrameNet* project.

- A representative sample of sentences has been taken from the *Health Science Corpus* and in collaboration with the *Spanish FrameNet* project (*FNE*), the automatic labelling of semantic roles with the SEMAFOR statistical tagger has been carried out. However, the semantic annotation was highly imprecise, due to the specificities of biomedical English. As it has been noted before, the domain-specific peculiarities of specialized languages usually undermine the reliability of NLP tools. Walter (2009), for example, has shown that the precision of the parser (PREDs) decreased to 64% when it was used to parse a corpus of court decisions, in contrast to a reliability of 86% when analyzing a corpus of newspapers.
- In this case study I have carried out an analysis of the semantic frames that a selection of the verbs included in *SciE-Lex* evoke and of their valences, following the descriptions provided by *FrameNet*. I have analyzed the concordance lines extracted from the *Health Science Corpus* and manually annotated the frame elements with their corresponding instantiations to check whether the information in the Berkeley *FrameNet* project is adequate for this specific register, that is to say:
  - i. whether a new frame has to be established or an existing frame customized,
  - ii. whether the meaning of the lemma in the *Health Science Corpus* suits the definition in *FrameNet*,
  - iii. whether all core semantic arguments can be described in terms of the frame elements in *FrameNet*.

Sentences have been annotated for the syntactic and semantic combinatorial possibilities of the lexical units that evoke the frame. The annotation includes: the frame evoked by the target lexical unit, the frame elements instantiated by the different constituents of the sentence, their grammatical function and syntactic phrase type.

- The occurrences which do not meet the descriptions in *FrameNet* are noted so as to make domain-specific customizations, such as the introduction of one or more frame elements to an existing frame or the creation of additional frames for specific uses encountered in the *Health Science Corpus*.
- These manually annotated sentences will be eventually used as a training corpus for the automatic tagging of the *Health Science Corpus*.
- In addition, I have also taken into account the types of collocates found in the data, since they may be key in distinguishing the different meanings of a polysemous word or the different lexical units of the same frame. Verbs in the frame *Evidence*, for example, usually collocate with Noun Phrase Objects that present a property or a process, which are often indicated by a term containing the suffixes *-(a)tion, -th, -ity...* (*differentiation, regulation, mutation, specificity*), which are morphologically and semantically related. On the other hand, the collocates that occur in the frame elements *MANNER* and *DEGREE* differ in accordance with differences in the meaning of the verbs. *Suggest*, which is a verb placed at one extreme of the continuum of tentativeness, collocates with *strongly* or *conclusively*, which reinforce the degree of certainty (*the data*

*strongly suggest...*), whereas these adverbs do not collocate with the verbs in which certainty is already implicit in their meaning. (Verdaguer and Noguchi 2018). *FrameNet* illustrates the most typical collocates of a lexical unit in the annotated sentences which are shown but does not explicitly mark them.

- This analysis and annotation will be carried out for all the frame-evoking verbs in *SciE-Lex* so as to establish semantic networks. On the one hand, this thorough analysis of the different frames evoked by a verb will be a good complement to draw attention to the close links between the syntactic and semantic behaviour of words and highlight the syntactic and combinatorial differences already included in *SciE-Lex*. On the other, the network established by the lexical units belonging to the same frame will reflect their shared features, while the differences in their selectional preferences will also reflect their semantic differences.

## 5. Case study

### 5.1. Frame Topic

In this paper I will study the lexical units that belong to a frame which is largely present in the type of register that I am analyzing, *Topic*. In addition to this intraframe analysis, I will also take an interframe perspective and will focus on a polysemous lexical unit belonging to this frame, *treat*, which occurs with different meanings in the biomedical register, in order to create a network of semantic interrelationships.

The *Health Science Corpus* consists of written texts on biomedicine, thus, as texts concerned with reporting and discussing research, there is frequent occurrence of lexical units evoking the frame *Topic*, defined in the *FrameNet* project as: “A TEXT that a COMMUNICATOR produces has a TOPIC that it is about”.

Its core frame elements are:

- the COMMUNICATOR, “the person that has produced a TEXT on a TOPIC”:

(13) *We have addressed this issue by examining oxygen activation*

- the TOPIC or “what the TEXT is about”:

(14) *The nature of this process is addressed in the following section*

- and the TEXT, “a set of propositions that is coherent in being about a TOPIC”:

(15) *The present study answered two unanswered questions*

Non-core elements are DEGREE (*This possibility has not been **thoroughly** addressed*), MANNER (*X et al. 2017 **directly** addressed the role of one component*), PLACE or TIME.

The lexical units that evoke this frame are the verbs *address*, *concern*, *cover*, *discuss*, *dwell on*, *refer*, *regard*, *treat*, the nouns *regard*, *subject*, *theme* and *topic*, and the

prepositions *about*, *concerning* and *on*. However, as mentioned in the introduction, in this case study I will only cover the verbs. I will deal with them all, except for  *dwell (on)*, which does not occur in the *Health Science Corpus*, and *cover*, which does occur, but not evoking the frame *Topic*, but the frame *Protecting* or *Filling* (*The nerve was then covered by paraffin oil*). On the other hand, I will also analyze *deal (with)*, which, although not yet listed as belonging to this frame, also evokes it and is used in the definition of *address*.

### 5.1.1. *Address*

*Address* is defined in the *FrameNet* project as “Deal with a topic.” The analysis of this verb in the *Health Science Corpus* shows frequent explicit occurrences of the COMMUNICATOR as a human Noun Phrase Subject:

(16) *We also addressed the question of whether an increase...*

Or the TEXT as a non-human Subject:

(17) *Few studies have addressed the detailed molecular changes*

It is interesting to note, as I will discuss below, that the TEXT can be realized either by a Noun Phrase Subject or by an Adverbial (either an Adverb or Prepositional Phrase):

Adverbs: *elsewhere*, *below*

Prepositional Phrase: *by previous studies*, *in a previous study*, *In X et al*, *in future work...*

(18) *The nature of this process is addressed in the following section.*

The TOPIC is either a Noun Phrase Object (*problem*, *question*, *possibility*, *complexities*, *role*, *issue*, *effects*, *experiments*, *implications*), which usually occurs as a passive Subject:

(19) *Five main questions are addressed*

or a *wh*-clause:

(20) *We specifically addressed whether it is the mean length of telomeres...*

(21) *Whether this requirement reflects an essential action of dynein during mitosis has not been addressed previously.*

As for non-core frame elements, that of MANNER is especially relevant, since it elaborates on the issue which is communicated, explaining the way in which it has been carried out. It is realized either by adverbs (*directly*) or Prepositional Phrases (*by boiling the pellets*, *by examining the effects*, *by examination of...*):

(22) *We have addressed this issue by examining the effects of ...*

Other non-core frame elements also present in the corpus are DEGREE (*This possibility has not been **thoroughly** addressed*) and TIME (*This question is **currently** being addressed*).

### 5.1.2. *Concern*:

*Concern* has been defined in *FrameNet* as “Relate, be about.” This verb has a characteristic of its own, which differentiates it from other verbs in the same frame, since it is always used with the frame element TEXT (collocates such as *question*, *problem*) and not the COMMUNICATOR:

(23) *A **major research question** concerns the cellular roles performed by each motor*

It is worth noting that this frame element usually occurs with some kind of premodification highlighting it (*a major, a key...*). It is also to be noted that when a Human Noun Phrase Subject appears, *concern* is used in a different sense, and thus evokes another frame, Cause\_emotion (*We were concerned that the satellite may have an altered composition*).

However, the syntactic realizations of TOPIC are very similar to those of the verb *address*. It can be a Noun Phrase Object, with collocates belonging to the same lexical domain: *roles, detection, estimates*:

(24) *A second problem concerns **estimates** of expected rates of introgression*

a *wh*-clause:

(25) *A key question concerns **whether** active genes...*

or a PP [*with*]:

(26) *Studies of DNA replication have, so far, mainly been concerned **with the core reactions of synthesis***

As the above sentence shows, TIME frame elements (*so far*) are also present.

### 5.1.3. *Deal*

Although *deal* is more frequently used evoking the frame Resolve\_problems, as in:

(27) *when we understand how organisms **deal with** this challenge*

which can clearly be identified by its co-occurrence with words such as *challenge*, *problems* or *difficulties* realizing the frame element PROBLEM, the *Health Science Corpus* has attestations of *deal* in the frame TOPIC occurring with:

the frame element COMMUNICATOR, realized by a Noun Phrase Subject or a PP[*by*] in a passive sentence:

(28) *This issue has been dealt with by previous investigators*

TEXT:

(29) *The foregoing discussion has dealt with the serine phosphorylation of STAT3 and STAT1a.*

and TOPIC:

(30) *This issue has been dealt with by previous investigators*

#### 5.1.4. Discuss

*Discuss* is defined as “Provide a discussion of a Topic.” Here again, the COMMUNICATOR is a NP referring to the authors, usually the pronoun *we*:

(31) *We discuss the implications of our results*

or a Prepositional Phrase introduced by *by* as the passive agent:

(32) *As discussed by X*

The TEXT is frequently realized as an Adverb (*above, below*) or a Prepositional Phrase (*in the text, in the next section*). However, it can be also found as the authors' work, which can be connected to the frame element COMMUNICATOR by means of metonymic transfer:

(33) *Estimates of gene density has been previously discussed (X 2015)*

The TOPIC is a Noun Phrase Object referring to the research itself or, more often, some aspect related to it (*advantages, implications, significance, possibility, issues, differences, analysis, role, model, point, mechanism*):

(34) *This possibility is discussed below*

As for the non-core frame elements, *discuss* occurs with DEGREE (*more fully, in detail*):

(35) *This is discussed more fully below*

MANNER (*in terms of, in the light of*)

(36) *This mechanism can now be discussed in terms of an oligosaccharide substance*

Or TIME, as the previous example shows (*now*).

### 5.1.5. Refer

*Refer* has been defined as "mention or allude to." This is a particularly interesting lexical unit, because it has several closely related meanings which belong to different frames, which are all present in the *Health Science Corpus*, with similar complementation patterns. Thus, *refer*, in addition to **Topic**, can evoke the following frames: **Referring\_by\_name** (*X1 and X2 refer to the same object in the document*) (*We refer to this allele as DaPc6*); **Sending** (*X is referring this finding to the Justice Department*) and **Reference\_text** (*Figure 6 refers to the three major chromosomes*).

In the frame **Topic**, again the **COMMUNICATOR** is a Human Noun Phrase Subject:

(37) *We refer to the seven-member AAD gene set*

The **TEXT**, as is also frequently the case with the other verbs, is often implicit. However, the most characteristic feature of *treat*, which sets it apart from the other verbs, is that the **TOPIC** is always introduced by a PP[*to*]:

(38) *PetI fragment, which is referred to above...*

The same complementation pattern occurs with this verb when it evokes other frames (as for example, *We refer to this determinant as a 'distributed' degron*: **Referring\_by\_name** or *Arrows refer to discrete bands*: **Reference\_text**), showing that the analysis in terms of frame semantics is highly fine-grained, since it can clearly distinguish several distinct meanings of a verb occurring in the same syntactic context.

### 6.1.6 Treat

*Treat* has been defined in FrameNet as "deal with some topic." In the sentences with *treat* the **COMMUNICATOR** is realized by a human Noun Phrase or is often left implicit (**CNI**) in passive sentences. As for the **TEXT**, it is also often left implicit:

(39) *Estimates need to be treated with caution*

The **COMMUNICATOR**, the author(s) of the article is the implicit passive agent, which is a constructional null instantiation, licensed by the grammar of English, whereas the **TEXT**, the article itself, is the definite null instantiation.

The **TOPIC** is again expressed by means of a Noun Phrase Object or a passive Subject:

(40) *These intronic results should be treated with caution*

As for the non-core frame elements, in the corpus there are occurrences of **MANNER**:

(41) *Estimates need to be treated with caution*

And EXPLANATION:

(42) *They should be treated with caution because of the small sample size.*

Other verbs, such as *discuss* or *address* are more typically used in this sense, so there is a greater variety of realizations of the different frame elements, as we will see in 6.1.7.

### 5.1.7. Summary of the results and discussion

Verbs	COMMUNICATOR	TEXT	TOPIC
<i>address</i>	NP Ext Human	NP Ext Adv PP	NP Obj wh-cl
<i>concern</i>		NP Ext	NP Obj wh-cl PP
<i>deal</i>	NP Ext Human	NP Ext	PP
<i>discuss</i>	NP Ext Human	NP Ext Adv	NP Obj wh-cl
<i>refer</i>	NP Ext Human		PP
<i>treat</i>	NP Ext Human	NP Ext Adv	NP Obj

Table III. Core frame elements in verbs evoking the frame *Topic*

The semantic and syntactic description of lexical items in terms of frame semantics allows the categorization of lexical items according to the frames they evoke. The frame semantics approach facilitates the establishment of the interrelationships of the words belonging to the same semantic frame and the identification of the distinct characteristics underlying the usage of the individual lexical units. This theoretical framework allows us to systematically analyze the polysemous structure of lexical items and at the same time to integrate the description of the meaning of individual words into a higher level of lexical organization in order to highlight the interconnections among the lexical units that evoke the same frame.

The verbs that belong to the frame *Topic* are closely related. However, there are also obvious differences among them, shown in the definitions and reflected in the salience of the different frame elements. Thus, two relevant features of the verbs in this frame are the occurrence or implicitness of two core frame elements, the *COMMUNICATOR* and the *TEXT*, and the metonymic connection between *TEXT* and *COMMUNICATOR*, reflecting a difference between the verbs with more personal involvement (*address*, *discuss*) and those with a lower degree of personal involvement (*concern*).



All the verbs evoking the frame TOPIC except *concern* can occur with a Noun Phrase Subject which can be human or non-human. When it is human, it occurs as the frame element COMMUNICATOR. If it is non-human, it is the TEXT. In the case of *concern*, which has characteristics of its own, in the *Health Science Corpus* there are only examples of the Subject as a non-human Noun Phrase, the TEXT, not the COMMUNICATOR. In many sentences one of these two core frame elements, COMMUNICATOR and TEXT, is implicit. Although both frame elements can occur, this is usually only the case when the TEXT is realized by a Prepositional Phrase or an Adverb in the function of an Adverbial.

In sentences with the verb *discuss* or *address* the subject is typically the COMMUNICATOR (*In this section I want to discuss the rather different possibility that some changes are essentially random* or *Here we have addressed two questions related to telomere length regulation*), whereas, as can be seen in the same examples, the TEXT is usually realized by a prepositional Phrase (*in this section, in the next section*) or an adverb (*here, above, below*). Only in a few cases the subject is non-human and instantiates the TEXT (*The study addressed two unanswered questions*). At the other end there is the verb *concern* with only the TEXT as the subject (*These questions concern rhetorical issues*).

Among the authors' strategies to depersonalize the article, in addition to passive structures, where the COMMUNICATOR is left implicit as a constructional null instantiation and the TOPIC is placed in thematic position (*This mechanism can now be discussed in terms of an oligosaccharide substance*), there is the TEXT occurring as the subject of the sentence. This can then be considered a metonymic extension, the product of a human activity (a TEXT) being used instead of a human subject:

- (43) *We discuss our results below*
- (44) (*X, unpublished observation*)
- (45) *As discussed by X et al. 2016*
- (46) *Estimates of gene density per unit chromosome has been previously discussed (X 2015)*

*We* or the author's name clearly refer to a human subject (the COMMUNICATOR). When the name of the author or authors appears without a date, for example in (name of the author, *unpublished observation*), it still refers to the COMMUNICATOR. However, when the date is added (*by X et al. 2016*) then it refers to the authors' work, that is to say, the the TEXT.

The TOPIC is usually instantiated by a Noun Phrase Object or, if it is more complex, by a *wh*-clause, but in the sentences with the verb *treat*, and in contrast to the other lexical units belonging to the same frame, the TOPIC is always introduced by a PP [*to*]:

- (47) *PetI fragment, which is referred to above...*

With the verb *concern*, on the other hand, in addition to a Noun Phrase or a *wh*-clause, the TOPIC can be realized by a PP [*with*]

(48) *Studies are mainly concerned with the core reactions of synthesis*

A thorough and careful analysis of frame elements is needed, since some Prepositional Phrases can perform different roles. For example, a Propositional Phrase introduced by *by* can realize the core element COMMUNICATOR (passive agent) or the non-core element MANNER (when followed by a non-finite *-ing* clause):

(49) *We have addressed this issue by examining the effects of ...*

The presence and frequency of non-core frame elements must also be taken into account, since they may reveal subtle distinctions in the meaning of the verbs evoking the same frame, such as their differences with respect to the salience of the TEXT or the COMMUNICATOR. MANNER (*X directly addressed the role of one component*), together with DEGREE (*This possibility has not been thoroughly addressed*), is one of the most relevant non-core frame elements. Both are particularly frequent, except with the verb *concern*, which is the one which is more focused on the TEXT and more depersonalized. Those focused on the COMMUNICATOR, on the other hand, frequently express the manner and degree of the action or event. These frame elements can be realized either by adverbs (*thoroughly, directly*) or Prepositional Phrases. TIME can also be present (*This question is currently being addressed*).

## 5.2. The polysemous verb *treat*

After having dealt with the syntactic and semantic behaviour of the verbs in the *Health Science Corpus* which evoke the frame TOPIC, I will now extend my survey to analyze the characteristics of one of the verbs, *treat*, to highlight its characteristic behaviour in its multiple meanings.

The verb *treat* in the Berkeley FrameNet project evokes the following frames:

Cure, Topic, Giving, Processing\_materials,  
Communicate\_Categorization, Medical\_intervention,  
Treating\_and\_Mistreating.

After a preliminary exploration of the *Health Science Corpus* I have found out that there are no occurrences of *treat* evoking the Giving frame (as in *The delegates were treated to an authentic Indonesian dinner*) and there is only one occurrence of the verb evoking the frame Treating\_and\_Mistreating (*All rats were treated in accordance with the European Community guidelines*), which will be ignored because of its low frequency. As the difference between Cure and Medical\_intervention is only that Medical\_intervention deals with attempts to alleviate a medical condition, whereas Cure deals with situations in which the Medical condition has been cured, I will be dealing with them in a unified way as Medical\_intervention. This leaves the following frames: Processing\_materials, Communicate\_Categorization, Medical\_intervention and Topic.

### 5.2.1. *Treat* in Processing\_materials

By far (97% of the occurrences), the most frequent use of *treat* in the *Health Science Corpus* is similar to that of *Processing\_materials*. This frame is defined as follows: “An AGENT alters some MATERIAL in some useful way by means of some chemical or physical Alternant. Typically, this involves placing a reagent in contact with the MATERIAL, or applying heat, pressure, etc.”

Its core elements are:

- The AGENT, the person who applies the Process to the MATERIAL:

(50) *We treated cells with 5 nmol*

- the ALTERANT, which causes a change in the MATERIAL:

(51) *Cells were treated **with chemicals***

- and the MATERIAL, which is altered by the AGENT in some useful way:

(52) *She treated **the lumber** with waterproofing fairy dust*

In addition, there are non-core frame elements such as DURATION, PLACE, MANNER, PURPOSE, RESULT or TIME.

In the occurrences of the *Health Science Corpus* there is an important difference with respect to the frame in *FrameNet*. Whereas in this project, what is usually treated are materials, as the name of the frame well indicates, in the *Health Science Corpus* what is processed are in most cases not really materials but cells or substances. In what follows I will show the typical syntactic functions and collocates found in this corpus with the verb *treat*.

The element AGENT is realized by the Noun Phrase Subject, or if the sentence is passive —as is often the case in scientific writing and in most of the occurrences found— it is usually not realized, and so it is a constructional null instantiation or, on rare occasions, a Prepositional Phrase introduced by *by*.

(53) *We treated cells with cycloheximide for 5 hours*

(54) *Cultures were treated as described above (CNI)*

The element ALTERANT is realized by a Noun Phrase Object. The collocates that realize it can be grouped into different lexical sets. The most frequent ones are:

Animals and plants: *Animals, rats, clones, hybrids, plants, explants, seeds, seedlings, protoplants*

*Cells (the most frequent one), membranes, nuclei, plasma, cultures, bacteria, viruses, reagent*

Laboratory objects: *Slides, plates.*

The element MATERIAL is usually realized by a Prepositional Phrase introduced by *with*:

(55) *The cell lines were treated **with various anti-cancer drugs***

One relevant and frequent construction worth noticing is that of the past participle *treated* modified by the ALTERANT:

(56) *A portion of each reagent was **heat-treated** by boiling for 20 m.*

where *heat* indicates the entity, which causes a change in the MATERIAL. Such structure is also found premodifying the MATERIAL: *pheromone-treated cells, heat-treated cells, HU-treated cells, retrovirally-treated cells...*

In addition, *treat* frequently occurs with non-core elements such as:

MANNER (realized by a Prepositional Phrase introduced by *by* or *with* or a non-finite clause introduced by *as*):

(57) *Cells were treated **as described in Figure 1***

(58) *Control cells were mock-treated **by the addition of DMSO***

(59) *They should be treated **with caution***

A caveat is in place here, since the Prepositional Phrase introduced by *with* usually introduces the core element MATERIAL. However, as the sentence above shows, it can also introduce the element MANNER. In fact, *with caution* is a frequent lexical bundle.

TIME:

(60) *Plates were treated **for 2 to 3 hours** with white light*

or PURPOSE (realized by a to-inf clause):

(61) *Plates were treated for 2 hours with white light **to induce germination***

### 5.2.2. *Treat* in Communicate\_Categorization

When the verb *treat* evokes other frames, it has a different syntactic and collocational behaviour. In the frame Communicate\_Categorization a SPEAKER communicates a message stating an ITEM's membership in a CATEGORY. The core elements are:

CATEGORY: "the class of entities with characteristics that match those of the ITEM":

(62) *They are treated **as dominant markers***

the ITEM “the entity that the speaker portrays as belonging to a CATEGORY”:

(63) *Size, measured by tibia length, was treated as a covariate*

the MEDIUM, “the piece of text in which a Speaker communicates their categorization of the ITEM”

and the SPEAKER, “the individual that communicates the message concerning the CATEGORY of an ITEM”, which in the *Health Science Corpus* are usually constructional null instantiations.

Less than 10% of the total occurrences of *treat* evoke this frame, with the element CATEGORY introduced by an *as-PP*. The MEDIUM and SPEAKER are mostly implicit, and the ITEM is instantiated by a NP (*insertion, size, rate, fragments*).

### 5.2.3. *Treat in Medical\_intervention*

*Medical\_intervention* is defined as “Procedural or Medicine based interventions are used on a Patient to attempt to alleviate a Medical condition [...]” This frame differs from *CURE* in that this frame deals only with attempts to alleviate a Medical condition, whereas *CURE* deals with situations in which the Affliction or Medical condition has been cured. For our purposes, we have fused them into one frame.

The core frame elements are:

INTERVENTION, “a drug or procedure administered or performed in order to treat a Medical condition”

(64) *Patients may do substantially better if treated from birth **with parenteral thiamine***

MEDICAL\_CONDITION, “A holistic description of the medical state of the patient (or a part of the state of the patient)”:

(65) *One strategy of treating **these diseases** is to inhibit new vessel growth*

MEDICAL\_PROFESSIONAL, the individual or team that attempts to improve the medical condition of the Patient:

(66) ***We** have treated several other children*

and the RESULT, the consequence of the Intervention:

(67) *The recovery rate in adults treated [...] is **quite modest**.*

The non-core elements are:

The PATIENT, the individual that receives medical treatment:

(68) *A young girl treated with high doses of ethinyl estradiol*

the EXTENT or degree to which an INTERVENTION has affected the MEDICAL\_CONDITION or Symptoms; the FREQUENCY\_of\_SUCCESS and the SIDE\_EFFECTS, which are often not present.

#### 5.2.4. Summary and comparison of the results

Frames	Core Frame Elements	Syntactic realization
Processing_materials	AGENT ALTERANT MATERIAL	NP External Human NP – Object PP [with]
Communicate_Categorization	CATEGORY ITEM MEDIUM SPEAKER	PP [as] NP
Medical_intervention	INTERVENTION MEDICAL_CONDITION MEDICAL_PROFESSIONAL RESULT	PP [with] NP NP External Human
Topic	COMMUNICATOR TEXT TOPIC	NP External Human NP – Object

Table IV. Frames that are evoked by *treat* and their core frame elements

As stated above, frame semantics allows us to systematically analyze the polysemous structure of lexical items, since the different meanings of a word evoke different frames. Unusually, since the various meanings of polysemous items occur in different syntactic patterns, the different meanings of *treat* occur with similar complementation patterns. Semantically, however, they are very different.

In all the different frames evoked by the verb *treat* there is a Human Subject in different roles: AGENT in Processing\_materials; SPEAKER in Communicate\_Categorization; Medical Professional in Medical\_intervention and COMMUNICATOR in Topic. They can be realized by personal pronouns (*I, we*) or left implicit in passive sentences (CNI). Except for Topic, where the Subject of *treat* can be the TEXT, all the other frames need a human being in the syntactic role of external argument. This fact is in line with the different meanings

of the verb, which need a human subject, although, also in line with their distinct senses, the role of the subject is different.

The Noun Phrase Object also has very different roles, with different types of collocates functioning as their head: ALTERANT in *Processing\_materials* (with different types of collocates: *animals, and plants, cells, membranes, slide, plates*); ITEM (with collocates referring to elements which can be categorized, for example *size*) in *Communicate\_Categorization*; Medical Condition (especially *diseases*) in *Medical\_intervention* and (*results*) in *Topic*. Once again, there are different collocates of *treat* which instantiate the different frame elements and help to distinguish them. Collocates are thus again revealed crucial to differentiate the various meanings of the lexical items and their evoked frames when they have the same complementation pattern.

One relevant construction, which only occurs in the frame *Processing\_materials* is that of the past participle modified by the ALTERANT:

(69) *A portion of each reagent was **heat-treated** by boiling for 20 m.*

and in some cases, premodifying the MATERIAL, such as *pheromone-treated cells*.

It is worth noting that this construction also occurs with the morphologically related noun *treatment* in the frames *Processing\_materials* (*Heat treatment abolished all these effects*) and *Medical\_intervention* (*Estrogen treatment has little effect on body weight*). However, with the verb, it occurs only with *Processing\_materials*. Again, this is in line with the semantic features of the verb. In *Medical\_intervention* the abstract noun *treatment* can be premodified by the type of medical care (*estrogen treatment*), but this is not possible with the verb (*\*estrogen-treated patients*), since with a human object the type of medical care (Intervention) has to be specified by means of a Prepositional Phrase introduced by *with*. MATERIAL in *Processing\_materials*, on the other hand, can be instantiated by a past participle modifier or PP [with] (*Lymphoblasts were treated with AMD*).

The variety of roles instantiated by the same pattern may also be highly relevant for the semantic description of a word. PP [with], in addition to MATERIAL in *Processing\_materials* and INTERVENTION in *Medical\_intervention* can also realize the non-core frame element MANNER (*with caution*). The type of collocate used in the Prepositional Phrase will again differentiate the frame elements. A careful analysis is thus needed in order to distinguish the different roles of an identical pattern. On the other hand, a frame element can be realized by different patterns, as the non-core frame element MANNER can be also instantiated by adverbs or a non-finite clause introduced by *as* (*the cell samples were treated as described above*). However, when *as* introduces a Prepositional Phrase, this instantiates the core frame element CATEGORY in the frame *Communicate\_Categorization* (*size was treated as a covariate*).

## 6. Conclusions and future work

This case study has illustrated the procedure to be followed in large-scale research. First, the differences among the frames in the Berkeley *FrameNet* project and in the *Health Science Corpus* need to be addressed. This is an obvious fact already shown by the *FrameNet* projects dealing with specialized languages. Although many frames are common, others have to be added or customized. In this exploration it has already been found that the frame `Processing_materials` needs to be adapted to the characteristics of the biomedical language.

Secondly, on the basis of the semantic frames which have been identified and on the evidence of the *Health Science Corpus*, the similarities and differences in meaning and syntactic patterning among the lexical units belonging to the same frame need to be highlighted. Differences in syntactic patterning or in the profiling of one frame element over the others may be highly relevant for the meaning of a word. Common features and subtle differences in meaning can explain differences in the salience of some core frame elements or the presence of non-core elements.

Thirdly, the similarities and differences among the different meanings of polysemous words need to be stressed in terms of the frames they evoke. A frame-based analysis has proved to be highly fine-grained, since it can distinguish the closely related meanings with the same syntactic realization. Syntactic patterning does not always distinguish the different meanings of a polysemous word. This can be illustrated by the verb *refer* that is always followed by a PP [to], regardless of the frame it evokes. It is also important to enhance semantic descriptions, including the collocational preferences of the lexical units. Although *FrameNet* does not explicitly deal with collocational patterns, I believe, as Johnson and Lenci argue (2013:26) "that the semantic description of LUs would be greatly enhanced by integrating this information on their selectional preferences in the *FrameNet* database." Collocates may also be crucial to distinguish the different meanings of polysemous words and these selectional preferences are approached in terms of semantic frames, as the establishment of the relations among lexical units in terms of frames gives systematicity to the analysis of the collocates.

Finally, networks of meaning will be established not only within frames –in fact, frames by themselves constitute a network of meaning as they characterize semantic relations between words– but among the frames that are evoked by polysemous words. These interframe analyses will reveal the closeness or distance of their different senses, which will be explicitly related to their syntactic and collocational patterns and will capture the interconnections and semantic closeness among words sharing more than one frame. Some of the lexical units evoking the frame `Topic` have been found, in addition, to share other frames. Thus, for example, *address* and *deal (with)* also evoke the frame `Resolve_problem`, so they show a higher degree of semantic closeness, which will be displayed in the semantic network.

Future research also needs to examine frame-bearing nouns, since nouns phrases and nominalisations are characteristic features of scientific writing (Salager-Meyer 1985;



Horsella and Pérez, 1991). Although it is usually the sentence main verb the lexical unit that evokes a semantic frame, the dominant semantic frame of a sentence can also be evoked by a noun. In this type of sentences, the noun refers to an event or state and has its own frame elements. The support verb, which has bleached meaning, is usually selected by the noun. Thus, for example in the sentence

(70) *Another member of the laboratory staff received treatment for a conjunctivitis*

is equivalent to:

(71) *Another member of the laboratory staff was treated for a conjunctivitis*

since they report on the same event. In (70) the frame-bearing word is *treatment*, not the verb, which is the support verb. The frame elements are: PATIENT (*Another member of the laboratory staff*) and AFFLICTION (*conjunctivitis*) and these Noun Phrases instantiate the same frame elements in both sentences. It has to be mentioned that the definition of support verb in *FrameNet* is broader than what is usually understood by support verbs, since they are described in the project as "verbs that turn a target noun (event or state) into a verb-phrase-like predicate, allow for the expression of a frame element as their subject, and are semantically neutral." So, nouns and their associated support verbs, which may distinguish the different senses of a word, also need to be addressed.

The implications of this corpus and frame-based theoretical study are lexicographic and pedagogic. *SciE-Lex* will be customized to include this new information based on frame semantics. With the implementation of *FrameNet*, *Scie-Lex* will be taken a step further to become a corpus-based and theoretically-driven database, not only providing a comprehensive and systematic description of the syntactic, semantic and combinatory behaviour of the language in the biomedical register, but also characterizing the semantic relations between words in terms of frames and the relations between those frames. Users will then find not only to know how to produce correct sentences but also the different alternatives that they have at their disposal to avoid repetition, use paraphrases and provide their texts with stylistic elegance. The user will be able to go from meaning to form and query how a particular frame element is syntactically expressed or go from form to meaning and search for the frame elements that are expressed by a particular construction.

Whereas the relevance of frame semantics in lexicography is well known (Atkins et al., 2003; Atkins and Rundell, 2008; Fontenelle, 2012; L'Homme, 2008, 2010, 2014; Kövecses and Csábi 2014), there have been few studies approaching the use this theoretical framework in teaching (Blanco 2006; Kövecses and Csábi 2014). However, there is great potentiality in using frame semantics once language teachers have acquired the necessary theoretical background knowledge to take advantage of the precision and systematicity of the syntactico-semantic information it can provide to prepare effective education materials (Verdaguer and Noguchi 2018). Frame semantics and its implementation in the project *FrameNet*, which systematize the relations between related lexical units and provide information on syntactic realization can be used by teachers in order to group the lexical items belonging to the same frame and

thus give coherence and systematicity to their teaching. As Kövecses and Csábi note (2014, p. 130) “awareness and acquisition of the cognitive structure of meanings aids vocabulary teaching and learning.” In the context of English for Specific Purposes or of CLIL (Content and Language Integrated Learning) (Muñoz 2007), where experts recommend systematic attention not only on the contents but also on the language development of the learner, the focus and systematization that frame semantics provides on the shared semantic background of a group of words, reflecting the speakers' understanding of their experience, can increase their knowledge and motivation. I believe this perspective of the biomedical register, based on frame semantics, can also be successfully applied in the teaching of English for Specific Purposes.

## References

- Atkins, Sue; Fillmore, Charles J.; Johnson, Christopher. R. 2003. Lexicographic relevance: Selecting information from corpus evidence. *International Journal of Lexicography* 16.3: 251-280.
- Atkins, Sue; Rundell, Michael; Sato, Hiroaki. 2003. The contribution of FrameNet to practical lexicography. *International Journal of Lexicography* 16.3: 333-357.
- Atkins, Sue; Rundell, Michael. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Biber, Douglas. 2006. *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam / Philadelphia: John Benjamins.
- Biber, Douglas. 2009. A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics* 14: 275-311.
- Biber, Douglas; Conrad, Susan; Cortes, Viviana. 2004. If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25.3: 371-405.
- Biber, Douglas; Barbieri, Federica. 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26: 263-286.
- Blanco, Olga. 2006. FrameNet as a corpus tool for the learning of second languages and for the lexical awareness of one's first language. *Porta Linguarum* 6: 67-76.
- Boas, Hans C. 2009. *Multilingual FrameNets in Computational Lexicography: Methods and Applications*. Berlin/New York: Mouton de Gruyter
- Borin, Lars; Dannélls, Dana; Forsberg, Markus; Gronostaj, Maria Toporowska; Kokkinakis, Dimitrios, eds. 2010. The past meets the present in the Swedish FrameNet. Extended version of *EURALEX 2010* presentation. Accessed 2018.08.25. URL: [https://svn.spraakdata.gu.se/sb/fnplusplus/pub/SweFN\\_Euralex\\_extended.pdf](https://svn.spraakdata.gu.se/sb/fnplusplus/pub/SweFN_Euralex_extended.pdf)

- Burchardt, Aljoscha; Erk, Katrin; Frank, Annete; Kowalski, Andrea; Padó, Sebastian; Pinkal, Manfred, eds. 2006. The salsa corpus: a German corpus resource for lexical semantics. *Proceedings of LREC- 2006*.
- Carrió-Pastor, M. Luisa. 2017. Verbal phraseology: An analysis of cognitive verbs in linguistics, engineering and medicine academic papers. In R. Mitkov, ed. *Cognitive and Cultural Aspects of Phraseology*. Springer International Publishing, pp. 325–336.
- Cortes, Viviana. 2004. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23.4: 397-423.
- Coxhead, Averil. 2000. A new academic word list. *TESOL Quarterly* 34.2: 213-238.
- Cruse, David A. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Dolbey, Andrew. 2009. *BioFrameNet: A FrameNet Extension to the Domain of Molecular Biology*. PhD. Dissertation, University of California Berkeley.
- Evans, Virginia. 1998. *Successful Writing Proficiency*. Swansea: Express Publishing.
- Faber, Pamela. 2015. Frames as a framework for terminology. In H.J. Kockaert and F. Steurs, eds. *Handbook of Terminology*. Amsterdam / Philadelphia: John Benjamins Publishing Company, pp. 14-33.
- Faber, Pamela; Buendía Castro, Miriam. 2014. EcoLexicon. In C.V. Andrea Abel and N.Ralli, eds. *Proceedings of the XVI EURALEX International Congress* Bolzano: EURALEX, pp. 601-607.
- Faber, Pamela; León-Araúz, Pilar; Reimerink, Arianne. 2016. EcoLexicon: new features and challenges. In I. Kernerman, I. Kosem Trojina, S. Krek, and L. Trap-Jensen, eds. *GLOBALEX 2016: Lexicographic Resources for Human Language Technology* in conjunction with the 10th edition of the Language Resources and Evaluation Conference. Portorož, pp.73-80.
- Fillmore, Charles J. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech* 280: 20-32.
- Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6: 222-254.
- Fillmore, Charles J.; Petruck, Miriam R. 2003. Glossary. Background to FrameNet. *International Journal of Lexicography* 16.3: 359-361.
- Fillmore, Charles J.; Johnson, Christopher R.; Petruck, Miriam. R. 2003. Background to FrameNet. *International Journal of Lexicography* 16.3: 235-250.

- Fillmore, Charles J.; Baker, Collin. 2010. A frame approach to semantic analysis. In B. Heine and H. Narrog, eds. *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, pp. 313-339.
- Fontenelle, Thierry. 2012. WordNet, FrameNet and other semantic networks in the International Journal of Lexicography – The Net result? *International Journal of Lexicography* 25.4: 437-449.
- Granger, Sylviane; Meunier, Fanny, eds. 2008. *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins.
- Hasegawa, Yoko; Lee-Goldman, Russell; Kong, Albert; Akita, Kimi. 2013. FrameNet as a resource for paraphrase research. In M. Fried and K. Nikiforidou, eds. *Advances in Frame Semantics*. Amsterdam / Philadelphia: John Benjamins Publishing Company, pp. 109–132.
- Horsella, Maria; Pérez, Fresia. 1991. Nominal compounds in chemical English literature: Towards an approach to text typology. *English for Specific Purposes* 19.2: 125-139.
- Hunston, Susan; Francis, Gill. 2000. *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam / Philadelphia: John Benjamins.
- Hyland, Ken. 2008. As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27: 4-21.
- Johansson, Stig. 2011. Corpus, lexis, discourse: A tribute to John Sinclair. In T. Herbst, Faulhaber, S. and P. Uhrig, eds. *The Phraseological View of Language. A Tribute to John Sinclair*. Berlin: Mouton de Gruyter, pp. 17-26.
- Johnson, Martina; Lenci, Alessandro. 2013. Verbs of visual perception in Italian FrameNet. In M. Fried and K. Nikiforidou, eds. *Advances in Frame Semantics*. Amsterdam / Philadelphia: John Benjamins Publishing Company, pp. 13-50.
- Kövecses, Zoltán; Csábi, Szilvia. 2014. Lexicography and cognitive linguistics. *RESLA, Revista Española de Lingüística Aplicada* 27.1: 118-139.
- Laso, Natalia; John, Suganthi. 2013. An exploratory study of NNS medical writer's awareness of the collocational patterning of abstract nouns in medical discourse. *RESLA, Revista Española de Lingüística Aplicada* 26: 307–331.
- L'Homme, M. Claude. 2008. Le DiCoInfo. Méthodologie pour une nouvelle génération de dictionnaires spécialisés. *Traduire* 217: 78-103.
- L'Homme, M. Claude. 2010. Designing terminological dictionaries for learners based on lexical semantics: The representation of Actants. In P. Fuertes-Olivera, ed. *Specialised Dictionaries for Learners*. Berlin/New York: De Gruyter, pp. 141-153.

- L’Homme, M. Claude. 2014. Why lexical semantics is important for e-lexicography. *International Journal of Lexicography* 27.4: 360-377.
- L’Homme, M. Claude. 2018. Maintaining the balance between knowledge and the lexicon in terminology: a methodology based on Frame Semantics. In Peters, P., J. G. Yongwei and J. Ding, eds. *Medical Lexicography and Terminology*. Special issue of *Lexicography*. *Journal of Asialex* 4(1).
- L’Homme, M. Claude; Robichaud, Benoît. 2014. Frames and terminology: representing predicative units in the field of the environment. *Cognitive Aspects of the Lexicon* (Cogalex 2014), Dublin.
- L’Homme, M. Claude; Robichaud, Benoît; Subirats, Carlos. 2014. Discovering frames in specialized domains. *LRec*: 1364-1371.
- L’Homme, M. Claude; Subirats, Carlos; Robichaud, Benoît. 2016. A proposal for combining “general” and specialized frames. *CogAlex*, 156-165.
- Lowe, John; Baker, Collin; Fillmore, Charles J. 1997. A Frame-Semantic approach to semantic annotation. In *Tagging Text with Lexical Semantics: Why, What, and How?* Proceedings of the Workshop. Special Interest Group on the Lexicon. 18-24.
- Meunier, Fanny; Granger, Sylviane, eds. 2008. *Phraseology in Foreign Language Learning and Teaching*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Muñoz, Carmen. 2007. CLIL. Some thoughts on its psycholinguistic principles. *RESLA, Revista Española de Lingüística Aplicada* Volumen Monográfico: 17-26.
- Nattinger, James R; DeCarrico, Jeanette. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Paquot, Magali. 2010. *Academic Vocabulary in Learner Writing*. London / New York: Continuum.
- Reimerink, Arianne; Faber, Pamela. 2009. Ecollexicon: A frame-based knowledge base for the environment. In J. Hřebíček, J.H. Mírovský, W. Pillmann, I Holoubek and T. Bandholtz, eds. *European conference of the Czech Presidency of the Council of the EU TOWARDS eENVIRONMENT Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe* Brno: Masaryk University, pp. 25-27.
- Römer, Ute; Schulze, Reiner, eds. 2009. *Exploring the Lexis-Grammar Interface*. Amsterdam / Philadelphia: John Benjamins.
- Rundell, Michael et al., eds. 2007. *Macmillan English Dictionary for Advanced Learners* (2nd ed.). Oxford: Macmillan Education.

- Salager Meyer, Françoise. 1985. Syntax and semantics of compound Nominal Phrases in medical English literature: A comparative study with Spanish. *English for Specific Purposes*. (Special Issue: ESP in South America).
- Sato, Hiroaki. 2009. Browsing Japanese FrameNet with FrameSQL. *Proceedings of Pacific Association for Computational Linguistics (PACLING 2009)*, Sapporo.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, John. 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Schmidt, Thomas. 2009. The Kicktionary - a multilingual lexical resource of football language. In H. Boas, ed. *Multilingual FrameNets in Computational Lexicography. Methods and Applications*. Berlin / New York: Mouton de Gruyter, pp. 101-132.
- Stubbs, Michael. 2001. *Words and phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishers.
- Subirats, Carlos. 2009. Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon. In H. Boas, ed. *Multilingual FrameNets in Computational Lexicography. Methods and Applications* Berlin / New York: Mouton de Gruyter, pp. 135-162.
- Subirats, Carlos. 2013. Frames, constructions, and metaphors in Spanish FrameNet. In I. Verdaguer, N. J. Laso, and D. Salazar, eds. *Biomedical English: A corpus-based approach*. Amsterdam / Philadelphia: John Benjamins, pp. 185-210.
- Swales, John M. 1990. *Genre Analysis: English in Academia and Research Settings*. Cambridge: Cambridge University Press.
- Swales, John M. 2004. *Research Genres: Explorations and Applications*. Cambridge: Cambridge University Press.
- Venturi, Giulia. 2013. A semantic annotation of Italian legal texts. A FrameNet-based approach. In M. Fried and K. Nikiforidou, eds. *Advances in Frame Semantics*. Amsterdam / Philadelphia: John Benjamins Publishing Company, pp. 51-84.
- Verdaguer, Isabel; Laso, Natalia. 2006. Delexicalisation in a corpus of scientific English. In A. M. Hornero, M. J. Luzón, and S. Murillo, eds. *Corpus Linguistics. Applications for the Study of English*. Bern: Peter Lang, pp. 417-428.
- Verdaguer, Isabel; Laso, Natalia; Salazar, Danica, eds. 2013. *Biomedical English: A corpus-based approach*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

- Verdaguer, Isabel; Noguchi, Judy. 2018. Identifying verb collocational patterns in a specialized medical journal corpus: a pedagogical approach to phraseology. *RESLA, Revista Española de Lingüística Aplicada* 31.2: 660-685.
- Vincent, Benet. 2013. Investigating academic phraseology through combination of very frequent words: a methodological exploration. *Journal of English for Specific Purposes* 12: 44-56.
- Walter, Stephan. 2009. Definition extraction from court decisions using computational linguistic technology. In G. Grewendorf and M. Rathert, eds. *Formal linguistics and law*. Berlin and New York: Mouton de Gruyter, pp. 183-224.
- Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wray, Alison. 2008. *Formulaic Language: Pushing the boundaries*. Oxford University Press.