RAEL: Revista Electrónica de Lingüística Aplicada
Vol./Núm.: 20/1
Enero-diciembre 2021
Páginas: 71-93
Artículo recibido: 05/08/2021
Artículo aceptado: 09/12/2021
Artículo publicado 31/01/2022
Url: https://rael.aesla.org.es/index.php/RAEL/article/view/463

Disambiguating a Disambiguation Tool: *Babelfy* from a Linguistic Point of View

Desambiguando una herramienta desambiguadora: *Babelfy* desde un punto de vista lingüístico

NATALIA LÓPEZ-CORTÉS
UNIVERSIDAD DE ZARAGOZA

Babelfy is an online tool, developed in the context of Natural Language Processing. When an item with more than one meaning is introduced in *Babelfy*, it chooses the appropriate meaning considering the context. The objective of this research study is to test the Word Sense Disambiguation skills of *Babelfy* in Spanish from a linguistic approach. To do so, a descriptive and comparative study between *Babelfy* and native Spanish speakers was carried out. Twenty-two pairs of sentences with an ambiguous word were designed, the first sentence of the pair had a neutral context and the second one a facilitating context. These sentence-pairs were introduced in *Babelfy* to check which meaning of the ambiguous word was selected and to explore whether there were differences depending on the type of context. These results were then compared to the answers of sixty-two Spanish native speakers. The data show that the behaviour of speakers when encountering an ambiguous word is not equivalent to the way *Babelfy* performs Word Sense Disambiguation, especially when the context is neutral, and the word has related meanings.

Keywords: *disambiguation; ambiguity; context; NLP; polysemy*

Babelfy es una herramienta online, desarrollada en el marco del Procesamiento del Lenguaje Natural. Cuando un ítem con más de un significado se introduce, éste escoge el significado adecuado teniendo en cuenta el contexto. El objetivo de esta investigación es poner a prueba las habilidades de desambiguación de significados de *Babelfy* en español desde un punto de vista lingüístico. Para ello, se llevó a cabo un estudio descriptivo y comparativo entre *Babelfy* y hablantes nativos del español. Se diseñaron veintidós parejas de oraciones con palabras ambiguas, una con contexto neutro y otra con contexto facilitador. Estas parejas fueron introducidas en *Babelfy* para comprobar qué significado de la palabra ambigua era seleccionado y para observar si existían diferencias en la desambiguación dependiendo del tipo de contexto. Estos resultados se compararon con las respuestas de sesenta y dos hablantes nativos de español. Los datos muestran que el comportamiento de los hablantes al encontrarse con palabras ambiguas no es equivalente a la manera en que *Babelfy* produce la desambiguación, especialmente cuando el contexto es neutro y la palabra tiene significados relacionados.

Palabras clave: *desambiguación; ambigüedad; contexto; PLN; polisemia*

1. INTRODUCTION

Ambiguity is produced when a single string of sounds, letters or words has got more than one meaning. Depending on its origin, three different types of ambiguity can be found, as seen in examples (1)-(3):

(1) a. /ðeə/ > their/they're
b. /'tu βo/ > *tuvo/tubo*
'had/tube'

(2) a. bank > institution/of a river
b. *banco > institución/asiento/de peces*
'bank, bench, shoal > institution/seat/of fish'

(3) a. I saw the man with the binoculars
[I saw [the man][with the binoculars]]
[I saw [the man with the binoculars]]

b. *Hablé de mis pesadillas con la doctora*
'I talked about my nightmares with the doctor'
[*Hablé* [*de mis pesadillas*][*con la doctora*]]
[*Hablé* [*de mis pesadillas con la doctora*]]

English and Spanish cases of phonological ambiguity are presented in (1), where a single string of phonemes is linked with two different lexical units with different meanings. In (2) we find a single word with two (or more) meanings, which is an example of semantic ambiguity. Lastly, syntactic ambiguity is produced when a single sentence can project two different structures, as it occurs in (3).

Semantic ambiguity is the purest type of ambiguity from a linguistic point of view: it appears as a consequence of the word itself, and not because of how sounds or phrases are combined. Hence, an ambiguous word may contain several different meanings under one single lexical form, which challenges speakers and their interpretation of this type of lexical units, especially considering that semantic ambiguity is ubiquitous in all languages and more common than monosemy (i.e., the one-to-one mapping between form and meaning).

It is important to note that semantic ambiguity is not a homogenous phenomenon: there are two types depending on the relation between the different meanings:

(4) a. bark > of a dog/of a tree
b. *cardenal > de la iglesia/herida*
'cardinal, bruise > of the church/wound'

(5) a. rabbit > animal/meat
b. *conejo > animal/carne*
'rabbit > animal/meat'

In (4) we can observe how homonymy is produced when the different meanings of a word are not related, whereas polysemy takes place when the different meanings of a word are somehow related (5). For instance, polysemy can be produced when the primitive meaning has been extended through a process of metonymy, as shown in (5). Some authors (Pustejovsky, 1995 and, more recently, Youn, Sutton, Smith, Moore, Wilkins, Maddieson & Bhattacharya,

2016) consider that there are some structures prone to polysemy, coined as *regular polysemy*. These authors claim that structures such as 'animal' and 'meat-of-that-animal' are common in all languages and that there is some sort of universal procedure of meaning extension.

Semantic ambiguity types have traditionally been defined from a historical, diachronic approach; that is, considering the etymological history of words and its meanings. From this perspective, polysemy is produced when an ambiguous word has a single etymological origin, and homonymy when an ambiguous word has different etymological origins. However, in the last decades, subjective-approaches to polysemy and homonymy have been gaining ground. As stated above, these approaches are based on the relation between meanings and are the ones followed in the present work.

It has been argued that semantic ambiguity poses a challenge not only for speakers but also for machines and Artificial Intelligence (AI). The fact that a word can convey different meanings has strong implications in the field of Natural Language Processing (NLP). For instance, one of the major linguistic challenges of NLP refers to the complex task of programming a tool that can automatically select the right meaning out of the many options embedded in lexical units.

Recently, there have been some major efforts to design tools to disambiguate ambiguous units. The objective of the present study is to test one of these softwares from a linguistic point of view: *Babelfy* (Moro, Raganato & Navigli, 2014). To do so, this paper has been organised as follows: firstly, it gives a brief overview of how semantic ambiguity has been assessed in NLP, focusing on *Babelfy* software and its main characteristics. Then, a comparative study is presented to test whether *Babelfy* behaves in the same way as native speakers in order to disambiguate a sentence containing an ambiguous unit. Finally, we discuss the main differences between the disambiguation processes in order to establish what linguistic features should be taken into account to improve *Babelfy* and other similar tools.

2. SEMANTIC AMBIGUITY AND NLP

Jurafsky and Martin (2019: 1) say that "the idea of giving computers the ability to process human language is as old as the idea of computers themselves". NLP is "the subfield of computer science concerned with using computational techniques to learn, understand, and produce human language content" (Hirschberg & Manning, 2015: 261). Therefore, if NLP deals with human language, it needs to deal with one of its main central properties: ambiguity.

There has been a change of perspective when studying semantic ambiguity from a theoretical point of view. It used to be considered some sort of design flaw or even a handicap that speakers needed to overcome. However, ambiguity is proven to be a property of language that contributes to making language acquisition and the structuring of the mental lexicon more efficient (Solé, Coronimas-Murtra, Valverde & Steels, 2009; Srinivasan & Rabagliati, 2015). This quality of ambiguity has not only been perceived by linguists, but also by researchers in the NLP field (view, for instance, Schütze, 1997).

Even from a linguistic approach, the way speakers deal with ambiguity is almost a magical process, since speakers are able to communicate despite the highly ambiguous nature of language. We are often ignorant of words (or sentences) having more than one meaning in a common, daily conversation. Take, for instance, the sentence in (6) extracted from Wasow, Perfors and Beaver (2005: 273).

(6) Dogs must be carried.

The univocal interpretation would be the following: if you have a dog and you are in a place where there is a sign with the information in (6), it is compulsory to carry the dog in your arms. However, there is another possible interpretation if we only take into account the linguistic and semantic information presented in such utterance: in order to be in the place with this sign, it is compulsory to bring a dog with you. As Wasow et al. (2005: 273) say, "common-sense knowledge of the world keeps anyone but a linguist from noticing the former interpretation".

NLP tools are able, nonetheless, to obtain all the possible interpretation of an apparently simple and unambiguous sentence. A famous example, extracted from Pinker (2007: 209), shows all the possible meanings that a machine can find in the sentence *time flies like an arrow* in (7):

(7) a. Time proceeds as quickly as an arrow proceeds.
b. Measure the speed of flies in the same way that you measure the speed of an arrow.
c. Measure the speed of flies in the same way that an arrow measures the speed of flies.
d. Measure the speed of flies that resemble an arrow.
e. Flies of a particular kind, time-flies, are fond of an arrow.

All this data shows that approaching ambiguity (in general) and semantic ambiguity (in particular) in the field of NLP is a difficult, yet essential, task. The most common problem in Computational Linguistics is what is called the *granularity of senses*; in other words, the difficulty to establish how many senses a word has. Take, for instance, the example presented in (8), extracted from Schütze (1997: 68).

(8) a. The hunter brought a duck home.
b. The duck was swimming in the pond.

Does the word 'duck' have the same meaning in both sentences? One the one hand, it refers to the same external reference (an animal); on the other hand, the implied semantic features implied are not equal: in (8a) the duck is dead while in (8b) it is alive. Do we therefore have two different senses or are there some shades (or contextual variations) of the same, unique meaning? As Morton (1994: 81) claims, "no one has answered the question of how we may know with sharp clarity and definitive exactness when a word has one meaning alone and when it has two or more quite discrete meanings".

The problem of the granularity of senses can be approached from a pure linguistic perspective. To do so, it is important to establish a difference between the concepts of *meanings* and *senses*. In (4) and (5) examples of polysemy and homonymy were presented and, as it has been previously mentioned, the relation between the meanings of those types of units is not homogenous. It is more productive to consider that homonymous words have meanings, and polysemous words have senses (Rodd, Gaskell & Marslen-Wilson, 2002). That is, a homonymous unit has got two or more discrete, exclusive meanings and a polysemous word has got different senses that are connected to each other. Hence, the problem of granularity in NLP would only be extensive to polysemy, whereas the meanings of homonymous units, since they are discrete, could be considered as different, independent words.

In fact, homonymy is traditionally thought to be some sort of historical accident, produced by the convergence of two totally different units, with independent etymological origins, as is illustrated in (9):

(9) a. bark (of a dog) < from old English *beorcan*
bark (of a tree) < from proto-Germanic **barkuz*

b. *cardenal (de la iglesia)* < from Latin *cardinalis*
'cardinal (of the church)'
cardenal (herida) < from Spanish *cárdeno*
'bruise (wound)'

The idea that homonymy and polysemy are different phenomena has implications on how speakers access and process these types of units. In the last decades, psycholinguists have found that homonymy and polysemy show a different behaviour in online processing tasks. This points to a differential storage in the mental lexicon, defined as the set of mental representations of the units of our language in long-term memory that is accessed to determine meanings and understand and build linguistic expressions (Jacobs & Ziegler, 2015).

This separation between homonymy and polysemy is not a novel idea in linguistics and could be the answer not to solve but to reduce the problem of the granularity of senses. In NLP this has already been claimed. For instance, Schütze (1997) says the ambiguity presented in (10) is special, because there is no way to interpret (10a) as a mixture of the senses of (10b) and (10c). The reader needs to make a decision and pick one interpretation, whereas in (8) both readings (the duck is alive, and the duck is dead) may be possible with no further contextual information.

(10) a. I saw her duck.
b. I saw her bird.
c. I saw that she ducked.

Schütze (1997: 66) calls the ambiguity in (10) *Necker-cube ambiguity*, establishing an analogy with the shape presented in Figure 1. The idea is that the Necker-Cube (right) can be seen as a cube that is tilted downwards (middle) or upwards (left), but not both ways at the same time. Necker-cube ambiguity is a type of ambiguous unit that cannot be interpreted without selecting a meaning first. Thus, Necker-cube ambiguity represents homonymy.[1]

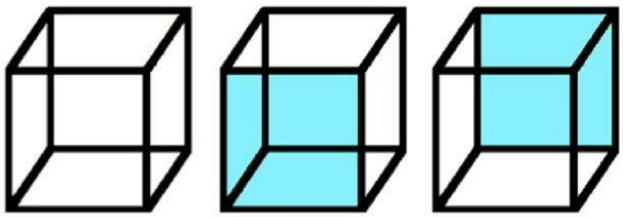

Figure 1: ***Necker-cube (left) and its two possible interpretations. Extracted from Ouhnana and Kingdom (2016: 60)***

Schütze (1997) then describes ambiguity as a cline, with vagueness and Necker-cube ambiguity (homonymy) at the extremes and polysemy at the middle. This is also a common idea in theoretical linguistics. For instance, when talking about semantic conceptualization in

[1] It is quite interesting to see how linguistic concepts are interpreted from a non-linguistic approach. This reinforces the idea that it would be very enriching to form multidisciplinary teams to solve problems related to NLP, such as the granularity of senses.

her theory of prototypes, Rosch (1975) says that belonging to one category or another is gradual and, when specifically studying ambiguity, Escandell (2008) claims that the relation between the meanings of a word is a question of degree. This necessarily connects with the idea that ambiguous units and their interpretation may vary depending on the speaker and his/her linguistic and extralinguistic knowledge (Haro, Ferré, Boada & Demestre, 2017). It would therefore be of great importance to consider the subjective interpretation of speakers when programming tools related to ambiguity.

Schtütze (1997) also presents ideas derived from psycholinguistics, as he introduces the importance of *coactivation* and claims that should be considered in NLP. Coactivation is produced when the different meanings of a word are activated regardless of the type of context used, even if this context facilitates one of the meanings among the others. This idea was formulated by Swinney (1979) and proved through an experimental design. Schütze (1997: 107) gives the example of the word 'image', presented in (11), to explain that in his hypothesis there is a type of ambiguity (polysemy) that is prone to coactivation.

(11) a. *Sense 1*. A picture formed in the mind.
b. *Sense 2*. The general opinion about a person, organization, etc., that has been formed or intentionally created in people's minds.
c. *Coactivated context*. The Hollywood senator had a noble looking image.

Schütze (1997: 70) believes that "in many cases, the speaker wants to use several meanings of an utterance". That is, in (11c) the speaker would mean both (11a) and (11b). However, we do not agree with this claim: from our point of view, speakers know what their utterances mean but sometimes the linguistic context speakers give does not help to produce a disambiguation.[2] Take, for instance, the two possible realizations of the Spanish phonetic scheme /'ba ka/, presented in (12).

(12) a. *vaca*
'cow'

b. *baca*
'roof rack'

When pronouncing the sounds, the speaker knows what meaning he or she wants to transmit. However, the hearer, without any contextual hints, may not be able to disambiguate. This is a phonetic ambiguity, as the one presented in (1) but we believe that the same process occurs with semantic ambiguity.

All in all, the main idea that Schütze (1997) and other researchers transmit is that any IA that wants to successfully deal with language (and ambiguity) needs to behave as close as possible to the way speakers do. That is why the comparative study presented in this paper can be the basis to enrich NLP tools. In this case, our objective is to study to what extent the behaviour of a tool varies compared to native speakers when disambiguating an ambiguous word in different contexts.

Another problem related to ambiguity and NLP is the fact that ambiguous words also vary depending on the language. As Jurafsky and Martin (2019: 358) say "the sense tag inventory for an English word might be the set of a different Spanish translation".

[2] This coactivated context is quite similar in structure to the neutral context used in our experimental design.

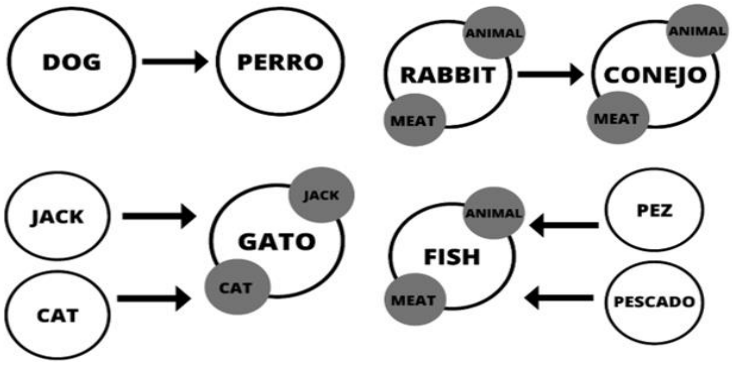

Figure 2: ***Ambiguous units and its different translations in Spanish and English. Adapted from López-Cortés (2019)***

In Figure 2, a comparison of ambiguous units between English and Spanish is displayed. There are different cases: a word with one meaning in one language is mapped with a word with one meaning in another ('dog' and *perro*); a word with multiple meanings in one language has the same multiple meanings in the other ('rabbit' and *conejo*); two different words in one language correspond to one single ambiguous word in another ('jack/cat' and *gato*) and one ambiguous word in one language corresponds to two different words in the other ('fish' and *pez/pescado*). This asymmetry between languages needs to be considered in NLP and is relevant for our research since *Babelfy*, the tool studied in the present paper, offers disambiguation processes for 271 languages.

Word Sense Disambiguation (WSD) is the "task of determining which sense of a word is being used in a particular context" (Jurasfky & Martin, 2019: 354). WSD is somehow a parallel process to human disambiguation. How speakers are able to disambiguate almost without being aware of the process is an interesting research question in linguistics and, to date, still unanswered:

> Is our talent for disambiguation a by-product of our general reasoning abilities, or did it develop in response to the ambiguity of language? If the latter, how did ambiguous languages emerge in the first place? (Wasow et al., 2005: 273)

A current biggest challenge of NLP is to achieve an adequate WSD. However, as Navigli (2009) claims, there are several problems regarding WSD, the major one being the multiple ways of understanding, defining and determining the discrete meanings of ambiguous units, as it has already been presented in this section.

Navigli (2009: 51) states that "WSD has not yet demonstrated real benefits in human language technology applications". Navigli (2009) lists applications of WSD such as machine translation or content analysis, among others. However, the problems of WSD are present in our daily life: for instance, users are aware of the limitations of search engines when something like what is captured in Figure 3 happens:

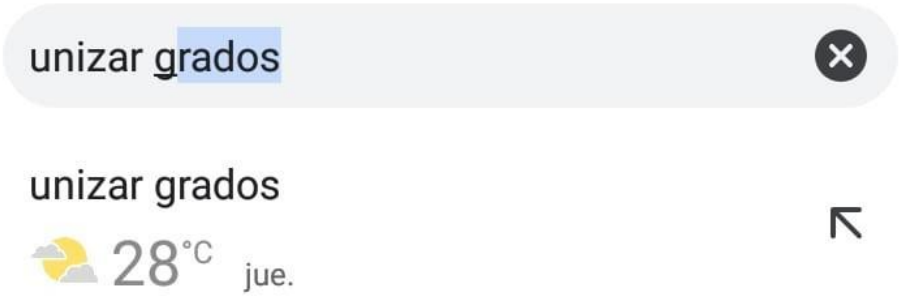

Figure 3: ***Search results of an ambiguous unit in Spanish***

Unizar grados was introduced in a search engine to obtain information about the academic degrees available at Unizar (Universidad de Zaragoza). However, the engine presented information about the temperature at Unizar, because *grados* is a Spanish ambiguous word: it can either mean *carrera universitaria* 'academic degree' or *grados centígrados* 'Celsius degrees'.

When a proper WSD procedure is obtained and mistakes like the one in Figure 3 are less common, NLP will be closer to achieving the Semantic Web (SW), which is "an extension of the current web, in which information is given well-defined meaning, better enabling computers and people to work in cooperation" (Berners-Lee, Hendler & Lassila, 2001). The main objective of this paper is to examine this problem with a linguistic approach, conducting a comparative analysis of *Babelfy* and native speakers' WSD skills.

According to Gale, Church and Yarowsky (1992), the WSD models have a performance of about 90% on average, against a human performance of 96-99%. However, as Navigli (2009) explains, an accuracy of above 95% can be obtained when disambiguating homonyms. The differentiation between homonymy and polysemy will therefore be considered in our research, since we believe it is one of the key aspects to improve WSD.

2.1 *Babelfy*

Within this context, *Babelfy* (Moro et al., 2014) is an online tool that disambiguates sentences. It is based on *BabelNet* (Navigli & Ponzetto, 2012), a semantic network that can be used as a source for lexical information from different languages.

Through BabelNet, users can access information regarding words and concepts, organized in so-called *synsets*. In Figure 4, the first four *synsets* of the word 'bark' are presented. As seen, it offers information about the grammatical category (in this case, noun, although there is also information about 'bark' as a verb in the search results) and the language.[3] It also shows a brief description and, sometimes, an image. The icons that are inserted in the image are related to the category of meaning: for instance, the tree means it is related to biology or the paw, to animals. The number under the icon shows the number of semantic connections the word has in the network.

[3] Taking a first look at this, some problems arise. Why is 'bark-of a dog' in the fourth place, below 'bark-a noise resembling the bark of a dog'? It is important to note that *BabelNet* works through an automatic process, gathering information from different sources (such as Wikipedia or WordNet). It is however curious because this problem somehow resembles the problem of circularity found in traditional dictionaries.

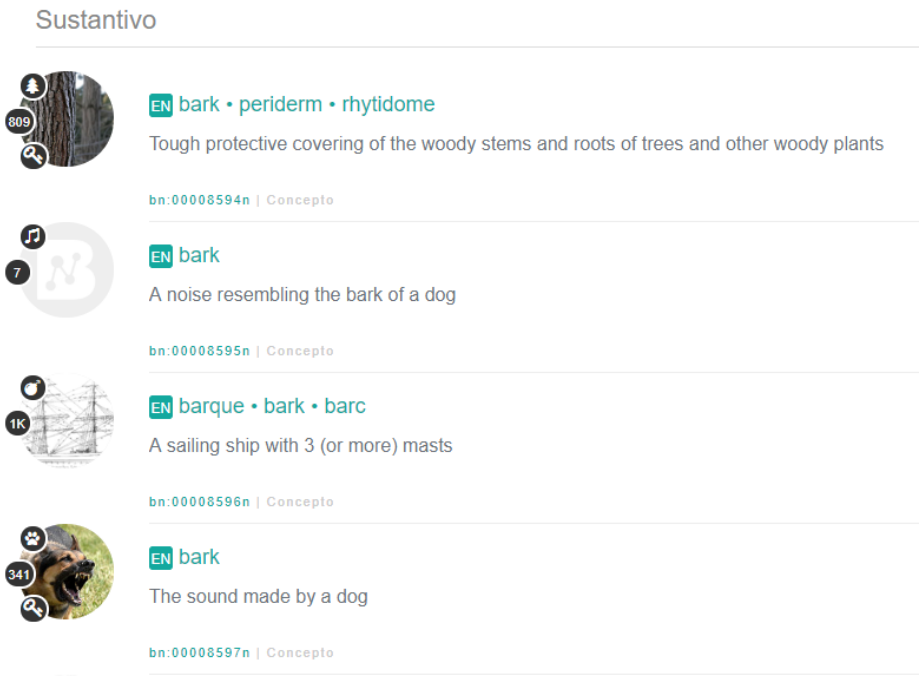

Figure 4: ***Sample of some of the results for the word 'bark' in BabelNet***

If one clicks on one of the meanings, further information can be accessed. Particularly interesting are the relations of the meaning, shown in Figure 5, that help to categorize the word and create its semantic network. This part of *BabelNet* is clearly linguistic: take, for instance, the 'part-of' or the 'is-a' relation, which are related to semantic relations such as hyperonymy or metonymy.

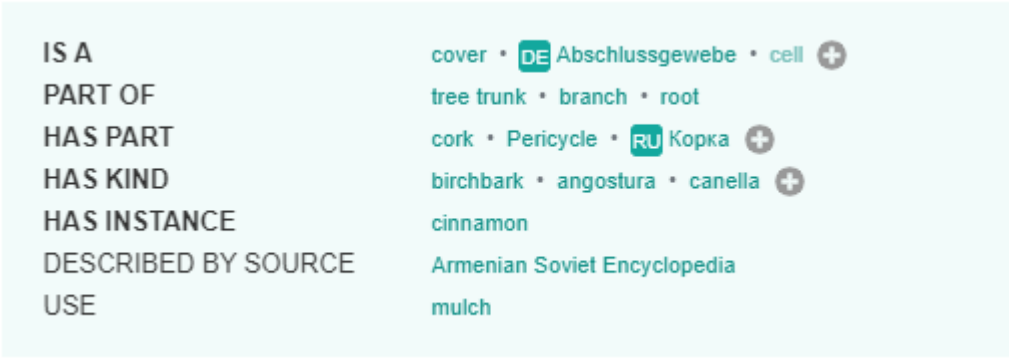

Figure 5: ***Further information of the synset 'bark-of a tree' in BabelNet***

BabelNet is available in different languages, which means one can obtain information about words from different languages, as Figure 6 shows with the ambiguous Spanish word *cardenal* ('cardinal' or 'bruise'). However, one of the most common meanings of this word, which is the meaning of 'bruise', is not among the first *synsets* shown in *BabelNet*. Although it derives from the meaning of *cardenal-color* ('colour'), which is in fact included, the meaning of 'wound' is far more common. That is why speakers' interpretations and subjective metrics should be somehow taken into account to make tools like this one truly useful.

Figure 6: ***Sample of some of the results for the word 'cardenal' in BabelNet***

Babelfy takes all this a step further since information about not only a word but a whole sentence can be gathered. It uses *BabelNet*'s *synsets* to link and disambiguate bigger pieces of lexical information. In Figure 7, a sentence is presented with all the pieces of information linked with their proper meanings.

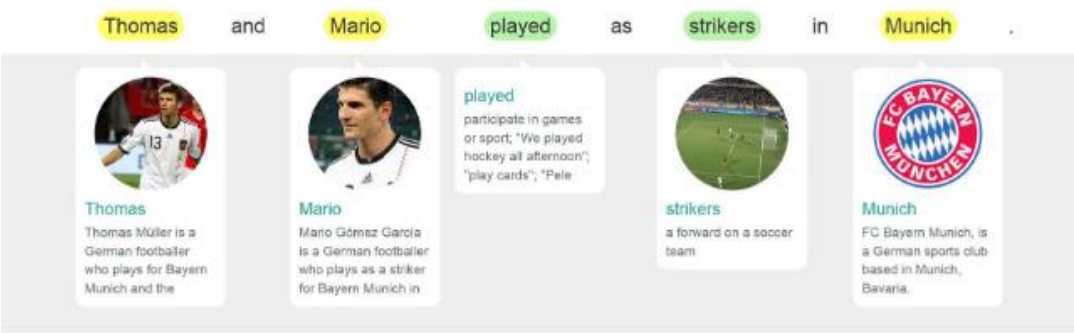

Figure 7: ***Example of a sentence in Babelfy. Extracted from Navigli (2015)***

As can be observed from Figure 7, *Babelfy* carries out different tasks, such as entity linking: *Babelfy* gives words' linguistic meanings and links words to their external references (as it occurs with the names 'Thomas' and 'Mario' or the football team 'Munich').

One of the most interesting and complex tasks that *Babelfy* carries out is WSD: that is, when introducing an item that has more than one meaning, *Babelfy* chooses the appropriate meaning (or *synset*) considering the context. For example, for the Spanish sentence *El gato se comió al ratón* ('The cat ate the mouse'), *Babelfy* selects the synset of 'mouse'-animal but for the sentence *El ordenador necesita un ratón* ('The computer needs a mouse'), it selects the meaning 'mouse-tool', as it can be seen in Figure 8:

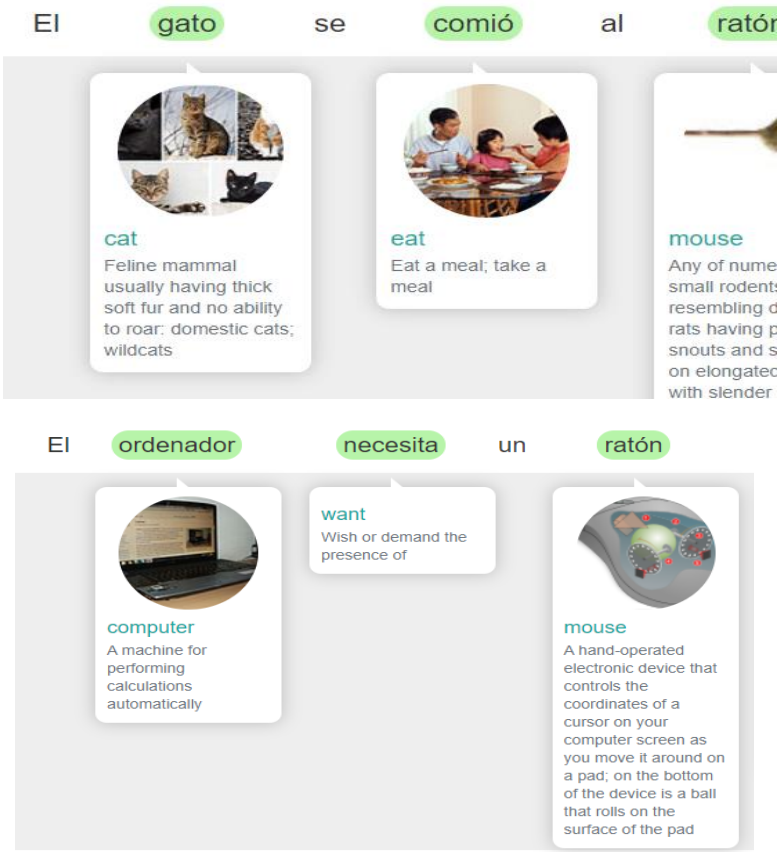

Figure 8: ***Two different sentences with the ambiguous word ratón ('mouse') in it***

3. SPEAKERS VS *BABELFY*: A COMPARATIVE STUDY

The present study tests how the *Babelfy*'s WSD works in Spanish depending on the context (facilitating or neutral) and on the type of ambiguity (homonymy or polysemy). In this section, the design of materials will be explained (3.1), followed by the procedure of the analysis (3.2) for both *Babelfy* and the speakers. Lastly, the results will be presented (3.3), as well as the consequent discussion (3.4).

3.1 *Design of materials*

The comparative work has got two variables of study: the type of ambiguity and the type of context. The material consisted of 22 pairs of Spanish sentences with an ambiguous word in them.

Of these 22 ambiguous words, 14 were homonymous and 8 were polysemous. They were classified according to subjective, psychological-relevant measures (see Haro, Ferré, Boada & Demestre, 2017 or López-Cortés, 2021). This distinction between types of ambiguity allowed us to study if there were any effects depending on the relation between meanings. Table 1 lists the ambiguous words used in the study, as well as their possible meanings:

Table 1: ***List of ambiguous units used in the comparative study***

HOMONYMY			POLYSEMY		
	Meaning 1	Meaning 2		Meaning 1	Meaning 2
Banco	*asiento* 'bench'	*de dinero* 'bank'	*Lengua*	*músculo* 'tongue'	*idioma* 'language'
Cardenal	*de la iglesia* 'cardinal'	*herida* 'bruise'	*Gallina*	*animal* 'chicken'	*cobarde* 'coward'
Heroína	*droga* 'heroine-drug'	*mujer* 'heroine-woman'	*Circulación*	*de la sangre* 'blood circulation'	*movimiento* 'movement'
Cabo	*accidente geográfico* 'cape'	*soldado* 'corporal'	*Busto*	*escultura* 'bust'	*pecho* 'breasts'
Caña	*de pescar* 'rod'	*bebida* 'beer'	*Canasta*	*recipiente* 'basket-container'	*de baloncesto* 'basket-of basketball'
Cubo	*figura geométrica* 'cube'	*recipiente* 'bucket'	*Armonía*	*paz* 'peace'	*musical* 'harmony'
Chorizo	*embutido* 'chorizo'	*ladrón* 'crook'	*Brote*	*de una planta* 'sprout'	*de una enfermedad* 'outbreak'
Estaciones	*de tren* 'train station'	*del año* 'season'	*Borrador*	*instrumento* 'rubber'	*papel* 'draft'
Golfo	*accidente geográfico* 'gulf'	*sinvergüenza* 'rascal'			
Pupila	*alumna* 'student'	*de los ojos* 'pupil'			
Servicio	*baño* 'lavatory'	*acción* 'service'			
Cólera	*ira* 'fury'	*enfermedad* 'cholera'			
Tiempo	*atmosférico* 'weather'	*cronológico* 'time'			
Pasta	*alimento* 'pasta'	*dinero* 'money'			

These ambiguous words were selected because their meanings were balanced: that is, there was not a prominent, more frequent meaning, which could be a flaw in the design of the materials. To determine whether these meanings were balanced, a questionnaire was carried out. In those questionnaires, 58 participants were asked to write all the meanings (of a list of words) they could think of. A word was considered to have two balanced meanings if they had a frequency of 70-30% among the answers. For instance, the word *estación* ('station'/'season') was accepted with the values of 'station' (65.51%) and 'season' (34.48%) but the word *lira* ('lyre'/'lira') was not, because the frequency was not among the limits (85% for the 'lyre' meaning and 15% for the 'lira' meaning).

Each balanced, ambiguous word was then inserted in a pair of sentences, one of which had a context that facilitated the disambiguation towards one of the meanings. The other sentence had a neutral context, valid for both meanings of the words. This is exemplified in Table 2, with a homonymous unit and a polysemous unit. The two sentences varied minimally: the change towards a facilitating context was made by introducing a phrase (as it can be seen in Table 2, where the changes in the facilitating context are underlined).

Table 2: ***Sample of the sentences used in the comparative study***

WORD	AMBIGUITY	NEUTRAL CONTEXT	FACILITATING CONTEXT	FACILITATED MEANING
Banco 'bank/ bench'	Homonymy	*Estuvo esperando durante horas. En el banco, sin moverse, pasó frio y por eso enfermó.* 'He waited for hours. In the *banco*, without moving, he got cold and that's why he got sick'	*Estuvo esperando sentado durante horas. En el banco, sin moverse, pasó frio y por eso enfermó.* 'He waited sitting there for hours. In the *banco*, without moving, he got cold and that's why he got sick'	*Banco-asiento* 'bench'
Borrador 'rubber/ draft'	Polysemy	*Se puso nervioso e hizo muchos cambios inútiles. Al final, dejó el borrador sobre la mesa.* 'He got nervous and made lots of pointless changes. Finally, he left the *borrador* on the table'.	*Se puso nervioso e hizo muchos cambios inútiles en el papel en sucio. Al final, dejó el borrador sobre la mesa.* 'He got nervous and made lots of pointless changes in the draft paper. Finally, he left the *borrador* on the table'	*Borrador-papel en sucio* 'draft paper'

The material for the comparative study consisted of 28 sentences with a homonymous word (14 with facilitating context and 14 with neutral context) and of 16 sentences with a polysemous unit (8 with facilitating context and 8 with neutral context). All these sentences were then disambiguated by *Babelfy* and by native Spanish speakers.

3.2 *Procedure*

The comparative study was conducted in order to test the WSD skills of *Babelfy*. We wanted to check whether there were any differences depending on (1) the type of ambiguity (homonymy-polysemy), (2) the type of context (neutral-facilitating) and (3) on the performance of *Babelfy* and native speakers.

Firstly, the 44 sentences were introduced in *Babelfy*, to check what meaning of the ambiguous word was selected. This way information was gathered about both neutral (Figure 9) and facilitating (Figure 10) contexts.

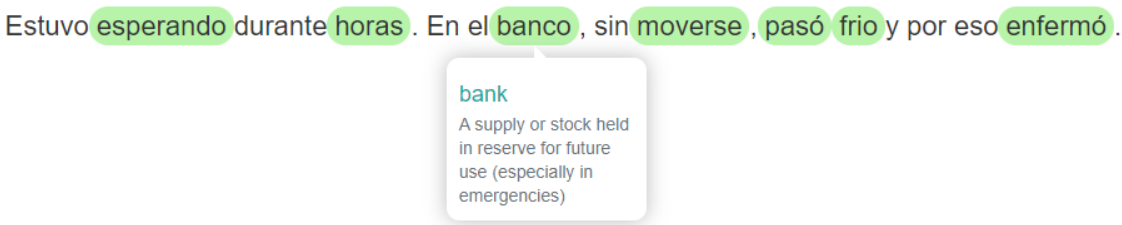

Figure 9: ***Babelfy results for the word banco with neutral context***

Figure 10: ***Babelfy results for the word banco with facilitating context***

In order to gather information on how native speakers disambiguate the material presented in the previous section, questionnaires were used. The questionnaires' structure is presented in Figure 11. The sentence was shown, followed by different options. These options consisted of possible meanings for the ambiguous word. In the example of Figure 11, the ambiguous word was *banco* ('bank/bench') and the possible answers were four: *banco-institución financiera* ('bank-financial institution'), *banco-mueble para sentarse* ('bench'), *banco-de peces* ('shoal') and *banco-orilla* ('bank-of a river').

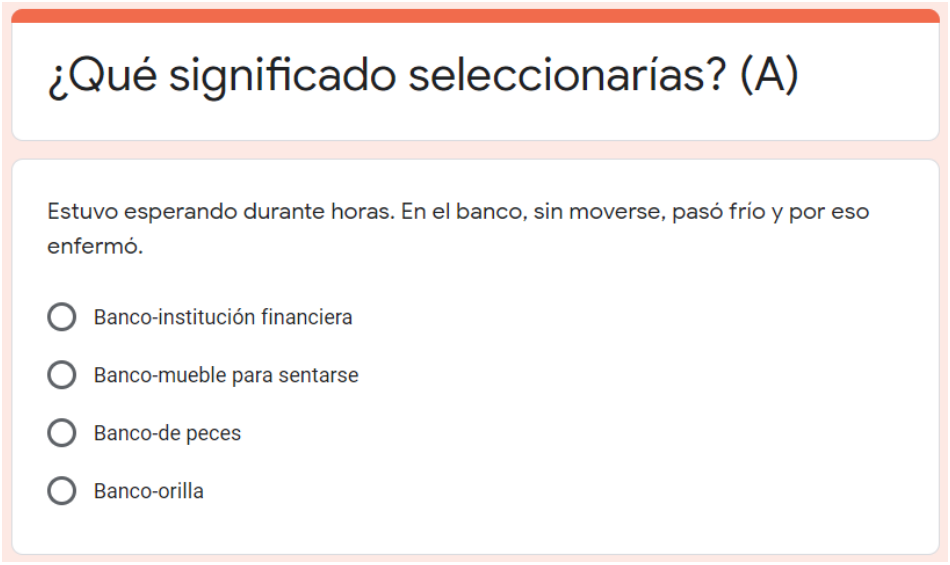

Figure 11: ***Sample of one of the questionnaire items***

These options were chosen because, as shown in Figure 12, they were the first four options available in *BabelNet*. By selecting the most common *synsets* of *BabelNet* we made sure that both the speakers and *Babelfy* had access to the same options for the disambiguation process.

Figure 12: ***First four synsets of BabelNet for the word 'banco'***

Two conditions were created and therefore two different questionnaires were designed. Each of them had 22 sentences, 11 with neutral context and 11 with facilitating context. This way the participants were exposed to each sentence only once and we were able to gather information about both contexts.

This study used a sample of 62 native Spanish speakers, who filled the questionnaires (31 participants per condition). All of them were aged between 18 and 25 and the questionnaires were completed in person.

At the beginning of the session, instructions were given orally. Participants were informed that their answers would be used statistically, and emphasis was made on the fact that there were no right or wrong answers: we wanted to know their interpretation as speakers. The session lasted around 10-15 minutes, depending on each participant's speed.

3.3. *Results*

In Table 3 the results for the condition 'neutral context' are presented. There are two groups of items (homonymy and polysemy) and for each of them different data are shown: the word followed by the disambiguation made by *Babelfy* and the native speakers. When *Babelfy* showed unexpected results (such as the meaning 'stock' for *banco*), a super-index was added. The link to that *synset* is presented below the table.

Regarding the speaker's data, there was variability in their questionnaire, so an agreement column is included in the table. For instance, all the participants chose the meaning of *cardenal-del la iglesia* ('cardinal'); however, only 64.5% of the participants chose the meaning of *heroína-droga* ('heroine-drug') instead of *heroína-mujer* ('heroine-woman').

Finally, both disambiguation processes can be compared, in terms of whether there was a match between the answers of *Babelfy* and the native speakers or not. As it can be seen in the table, when the context was neutral, there were 6 matches for homonymous words and 3 matches for polysemous units (a total of 9 matches). Regarding the mismatches, that is, the cases when *Babelfy* and speakers performed differently, there were 8 for homonymy and 5 for polysemy (a total of 13 mismatches).

Table 3: ***Results for the condition 'neutral context'***

	WORD	*BABELFY*	SPEAKERS	AGREEMENT	MATCH
Homony my	*Banco* ('bench/bank')	*almacén*[a] ('stock')	*asiento* ('bench')	93.5%	X
	Cardenal ('cardinal/bruise')	*de la iglesia* 'cardinal'	*de la iglesia* 'cardinal'	100%	✓
	Heroína ('drug/woman')	*héore* 'heroine-woman'	*droga* 'heroine-drug'	64.5%	X
	Cabo ('cape/corporal')	*accidente geográfico* 'cape'	*accidente geográfico* 'cape'	67.7%	✓
	Caña ('rod-beer')	*tallo* 'cane'	*bebida* 'beer'	77.4%	X
	Cubo ('cube/bucket')	*cantidad*[b] 'bucketful'	*figura geométrica* 'cube'	93.5%	X
	Chorizo ('chorizo/crook')	*embutido* 'chorizo'	*embutido* 'chorizo'	64.5%	✓
	Estaciones ('station/season')	*de tren* 'train station'	*de tren* 'train station'	58.1%	✓
	Golfo ('gulf/rascal')	*puta*[c] 'bitch'	*sinvergüenza* 'rascal'	71%	X
	Pupila ('student/pupil')	*de los ojos* 'pupil'	*de los ojos* 'pupil'	96.8%	✓
	Servicio ('lavatory/service')	*acción* 'service'	*acción* 'service'	61.3%	✓
	Cólera ('fury/cholera')	*enfermedad* 'cholera'	*ira* 'fury'	96.8%	X
	Tiempo ('weather/time')	*del reloj* 'time'	*atmosférico* 'weather'	80.6%	X
	Pasta ('pasta/money')	*pintura*[d] 'impasto'	*dinero* 'money'	96.8%	X
Polysemy	*Lengua* ('tongue/language')	*músculo* 'tongue'	*idioma* 'language'	77.4%	X
	Gallina ('chicken/coward')	*animal* 'chicken'	*cobarde* 'coward'	83.9%	X
	Circulación ('blood/movement')	*movimiento* 'movement'	*movimiento* 'movement'	90.3%	✓
	Busto ('bust/breasts')	*pecho* 'breasts'	*escultura* 'bust'	67.7%	X
	Canasta ('contaiter/basketball')	*juego*[e] 'canasta-game'	*de baloncesto* 'basket-of basketball'	90.3%	X
	Armonía ('peace-harmony')	*musical* 'harmony'	*paz* 'peace'	54.8%	X
	Brote ('sprout/outbreak')	*de una enfermedad* 'outbreak'	*de una enfermedad* 'outbreak'	83.9%	✓
	Borrador ('rubber/draft')	*papel* 'draft'	*papel* 'draft'	77.4%	✓

[a] See https://babelnet.org/synset?word=bn:00008367n&lang=EN&langTrans=EN
[b] See https://babelnet.org/synset?word=bn:00013591n&lang=EN&langTrans=EN
[c] See https://babelnet.org/synset?word=bn:00010740n&lang=EN&langTrans=EN
[d] See https://babelnet.org/synset?word=bn:00046078n&lang=EN&langTrans=EN
[e] See https://babelnet.org/synset?word=bn:00008885n&lang=EN&langTrans=EN

Table 4 shows the results for the condition 'facilitating context'. The structure of the table is the same as in Table 3, but in Table 4 a column presenting the facilitated meaning in the sentence has been added. As it can be seen in the table, this facilitated meaning and the answers

of the speakers were always the same.[4] There were 8 matches for homonymous words and 4 matches for polysemous units (a total of 12 matches). Regarding the mismatches, there were 6 for homonymy units and 4 for polysemy (a total of 10 mismatches).

Table 4: ***Results for the condition 'facilitating context'***

	WORD	FACILITATED MEANING	*BABELFY*	SPEAKERS	AGREEMENT	MATCH
Homonymy	*Banco* ('bench/bank')	*asiento* 'bench'	*institución* 'bank'	*asiento* 'bench'	100%	x
	Cardenal ('cardinal/bruise')	*de la iglesia* 'cardinal'	*de la iglesia* 'cardinal'	*de la iglesia* 'cardinal'	100%	✓
	Heroína ('drug/woman')	*droga* 'heroine-drug'	*droga* 'heroine-drug'	*droga* 'heroine-drug'	96.8%	✓
	Cabo ('cape/corporal')	*soldado* 'corporal'	*accidente geográfico* 'cape'	*soldado* 'corporal'	67.7%	x
	Caña ('rod-beer')	*bebida* 'beer'	*canción*[a] 'song'	*bebida* 'beer'	96.8%	x
	Cubo ('cube/bucket')	*recipiente* 'bucket'	*recipiente* 'bucket'	*recipiente* 'bucket'	64.5%	✓
	Chorizo ('chorizo/crook')	*embutido* 'chorizo'	*embutido* 'chorizo'	*embutido* 'chorizo'	87.1%	✓
	Estaciones ('station/season')	*del año* 'season'	*del año* 'season'	*del año* 'season'	100%	✓
	Golfo ('gulf/rascal')	*accidente geográfico* 'gulf'	*puta* 'bitch'	*accidente geográfico* 'gulf'	87.1%	x
	Pupila ('student/pupil')	*alumna* 'student'	*alumna* 'student'	*alumna* 'student'	54.8%	✓
	Servicio ('lavatory/service')	*acción* 'service'	*acción* 'service'	*acción* 'service'	96.8%	✓
	Cólera ('fury/cholera')	*ira* 'fury'	*enfermedad* 'cholera'	*ira* 'fury'	96.8%	x
	Tiempo ('weather/time')	*cronológico* 'time'	*cronológico* 'time'	*cronológico* 'time'	80.6%	✓
	Pasta ('pasta/money')	*pasta* 'pasta'	*pintura* 'impasto'	*dinero* 'money'	90.3%	x
Polysemy	*Lengua* ('tongue/language')	*idioma* 'language'	*músculo* 'tongue'	*idioma* 'language'	96.8%	x
	Gallina ('chicken/coward')	*cobarde* 'coward'	*animal* 'chicken'	*cobarde* 'coward'	100%	x
	Circulación ('blood/movement')	*movimiento* 'movement'	*de la sangre* 'blood'	*movimiento* 'movement'	100%	x
	Busto ('bust/breasts')	*pecho* 'breasts'	*pecho* 'breasts'	*pecho* 'breasts'	90.3%	✓
	Canasta ('container/basketball')	*recipiente* 'container'	*recipiente* 'container'	*recipiente* 'container'	93.5%	✓
	Armonía ('peace-harmony')	*musical* 'harmony'	*musical* 'harmony'	*musical* 'harmony'	58.1%	✓
	Brote ('sprout/outbreak')	*de una planta* 'sprout'	*de una enfermedad* 'outbreak'	*de una planta* 'sprout'	80.6%	x
	Borrador ('rubber/draft')	*papel* 'draft'	*papel* 'draft'	*papel* 'draft'	100%	✓

[a] See https://babelnet.org/synset?word=bn:08338353n&lang=EN&langTrans=EN

[4] There is one exception: *pasta*, where speakers interpreted another meaning ('money' instead of 'pasta'). This points to a flaw in the design of that facilitating context.

3.4 *Discussion*

The main goal of this comparative study was to test the WSD skills of *Babelfy*. Specifically, we wanted to check whether there were any differences depending on (i) the type of ambiguity (homonymy-polysemy), (ii) the type of context (neutral-facilitating) and (iii) the performance of *Babelfy* and native speakers.

Figure 13 shows the percentage of accuracy (%Acc, in shades of green) and the percentage of errors (%E, in shades of orange) of *Babelfy*, obtained by establishing a comparison with the speaker's performance. If the answers of *Babelfy* coincided with the ones made by the speakers, it was considered that *Babelfy* was performing accurately; if the answers were not equivalent, it was considered that *Babelfy* was making errors in the disambiguation process.

Therefore, the total accuracy of *Babelfy* when disambiguating ambiguous units in Spanish is 40.91% in a neutral context and 55% in a facilitating context, in comparison to the performance of Spanish native speakers. On the other hand, *Babelfy* produces 59.09% of errors when disambiguating ambiguous units in Spanish in a neutral context and 45.45% in a facilitating context.

In addition to the total percentage of accuracy and errors, Figure 13 shows the distribution of this data according to the type of ambiguity (homonymy and polysemy). When a word has got meanings that are not related (homonymy or Necker-cube ambiguity according to Schütze, 1997) *Babelfy* performs with 48.86% accuracy when the context is neutral and with 57.15% when it is facilitating (in contrast to 51.14% and 42.85% for errors). When a word has meanings that are related (polysemy) *Babelfy* performs with 37.50% of accuracy when the context is neutral and with 50% when the context is facilitating (in contrast to 62.50% and 50% for errors).

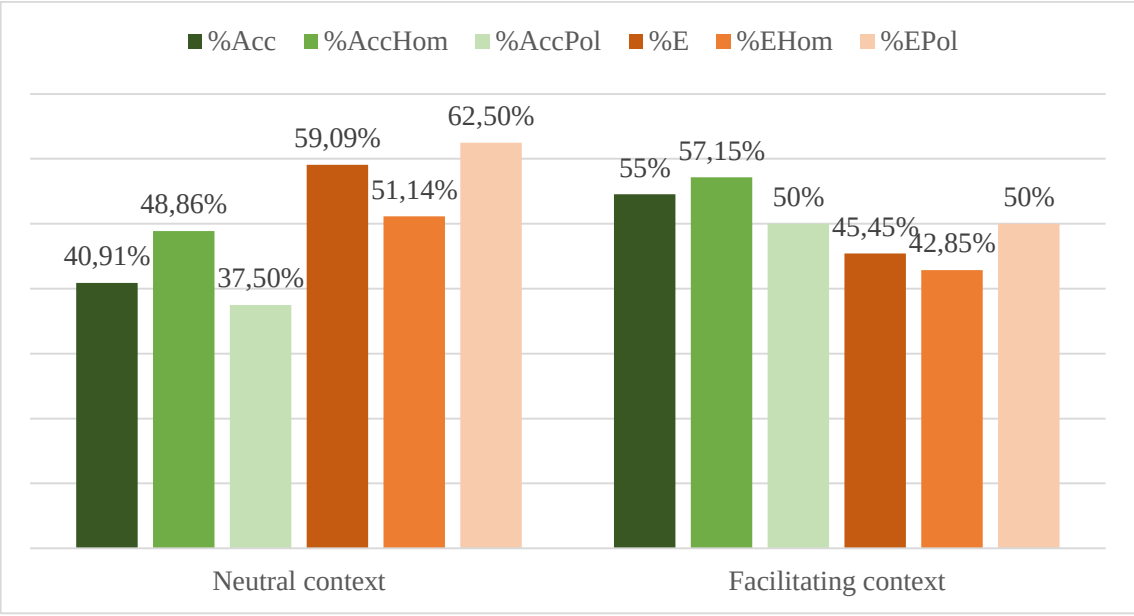

Figure 13: ***Summary of the results of the comparative study***

It therefore seems that *Babelfy* performs more accurately in Spanish, according to the native speaker standards, when the ambiguous unit is (1) embedded in a facilitating context and (2) is a case of homonymy. This makes sense if we consider that a homonymous form may convey different, exclusive meanings that need to be disambiguated to successfully process the

linguistic information. It is also coherent with the claim made by Navigli (2009) that homonymy can be more easily disambiguated that polysemy.

However, the results obtained in this study are contradictory to the prediction of the literature that the models of WSD can obtain an accuracy of above 90% (Gale, Church & Yarosky, 1992; Navigli, 2009). It is true that this is a modest study since the materials and the sample could be expanded; nevertheless, it is clear that we need to ask ourselves what "successful disambiguation" in NLP tools means. In this study, a successful disambiguation was considered to be a disambiguation equivalent to those made by native speakers. If this approximation is correct, it can give us clues to improve tools like *Babelfy*.

This is especially interesting when the context is neutral (Table 3), since speakers seem to use some disambiguating strategies that may be worth noting. For instance, there are cases where the agreement in choosing one meaning among the others is very high, or even total, as it occurs with the word *cardenal*: the meaning of *de la iglesia* ('cardinal') is chosen with a 100% of agreement even in the neutral context. That is probably because that meaning is more prominent in language use: the words of our material had balanced meanings, that is, participants were able to retrieve both meanings with a similar frequency. However, this does not mean that both meanings are used actively by the speakers in the same way: the linguistic knowledge (the semantic information that we store in our lexicon) and the use we make of that linguistic knowledge are not the same. When encountering a neutral context, without clues towards one meaning, it is not easy to know what the speaker means. Hence, it is natural to choose the meaning that we, as speakers, would use.

Furthermore, there is a tendency that seems to choose the more general term rather than the more specific one, as it occurs with the word *circulación* ('circulation'), that has a specific meaning (*de la sangr*e, 'blood') and a general meaning (*movement*, 'movimiento'). This could also point to collocations, that is, the linguistic phenomenon where two pieces of lexical information are tied together and most likely appear together (for instance *circulación sanguínea/de la sangre* in Spanish, 'blood circulation'). Therefore, if in the neutral context the word *circulación* appeared alone, it is more likely that it refers to the meaning of 'general movement'.

Yet, not all the information that speakers have is linguistic, and that is possibly the biggest problem that tools face. For example, in a neutral context, 98.6% of our participants chose the meaning of *cólera-enfermedad* ('cholera') instead of *cólera-ira* ('fury'), whereas *Babelfy* did the opposite. This is probably because speakers have extralinguistic knowledge related to the fact that cholera is no longer a common phenomenon, which means that its linguistic reference is less likely to be used.

The clues included in the sentence to produce the disambiguation were always perceived by the speakers and that is why they chose the meaning that was intended to be facilitated. As it can be seen in Table 4, *Babelfy*'s answers are not always equivalent to the speakers'. However, *Babelfy*'s performance can make sense because the disambiguation was not clear in all the cases (even though all the participants chose the expected meaning). Take, for instance, the pair of sentences in (13):

(13) a. *Pasaron varias horas perdidos en una niebla intensa. Finalmente, divisaron la silueta de un cabo solitario.*
'They were lost in a thick fog for hours. Finally, they saw the silhouette of a solitary *cabo*'

b. *Pasaron varias horas <u>en el campo de batalla</u> perdidos en una niebla intensa. Finalmente, divisaron la silueta de un cabo solitario.*

'They were in the battlefield, lost in a thick fog for hours. Finally, they saw the silhouette of a solitary *cabo*'

The ambiguous unit is *cabo*, which can either mean *accidente geográfico* ('cape') or *soldado* ('corporal'). The linguistic clue to facilitate the meaning of 'corporal' was to introduce the phrase *en el campo de batalla* ('in the battlefield'). Most of the speakers chose the expected meaning (almost 70%); however, *Babelfy* chose the meaning of 'cape'. This indeed makes sense because it could be totally plausible to be in a battlefield and find a cape. This implies that speakers do not only use linguistic clues to disambiguate, because the sentence in (13b) is also ambiguous and there is still a tendency towards one meaning (the same occurs with both the examples shown in Table 2). Discovering those extralinguistic strategies and trying to systematize them could be a step forward towards a better WSD.

There is also more information that speakers have in their lexicon and that *Babelfy* does not. This is the case, for instance, of diaphasic variation and colloquial register. In Spanish the word *caña* is widely used to refer to a pint of beer and in both contexts of our study a high number of participants chose this meaning over the other one (*caña* as cane). However, *Babelfy* does not recuperate the meaning of *caña*-'beer' in either case: in *BabelNet* the synset for this meaning is in ninth place, even though it is quite a common meaning in Spanish. A similar issue occurs with the meaning of *chorizo* as a robber instead of as food. Both meanings are present in *BabelNet*, as shown in Figure 14:

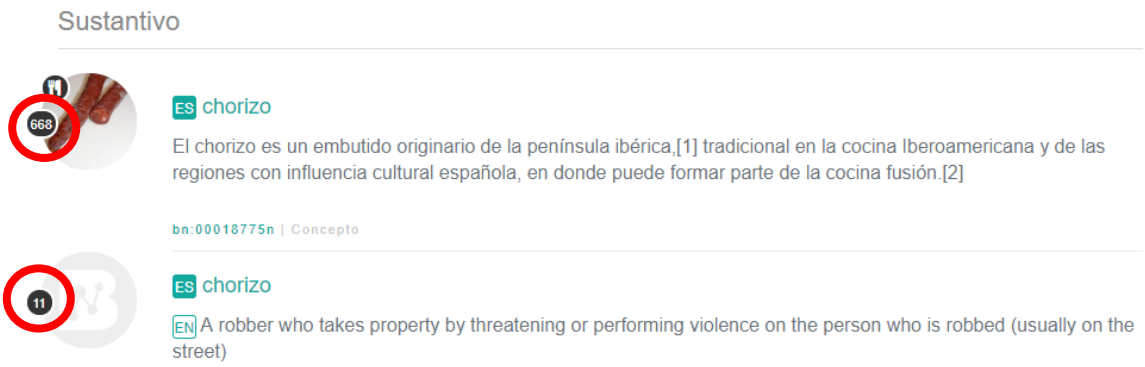

Figure 14: ***Sample of some of the results for the word 'chorizo in BabelNet***

However, there is a big difference between both meanings: *chorizo* as 'meat' has a total of 668 semantic connections in the database that enrich *BabelNet*, whereas the meaning of 'robber' only has 11 (this information is highlighted in Figure 14 using a red circle). If we go back to the synset of *caña* as 'beer', in ninth place of the list of possible meanings of the word in *BabelNet*, it only has 46 connections (versus more than 1,000 connections for the meaning of *caña* as 'cane'). This may indicate that the sources for *BabelNet* (and therefore *Babelfy)* are not heterogeneous and that getting access to more colloquial and every-day corpora may enrich the semantic networks that produce the WSD.

Finally, it is important to point out once again that *Babelfy* is a fantastic tool that conducts different complex processes and produces a highly efficient and automatic linking between linguistic units and possible meanings and, in the case of Entity linking, extralinguistic references. However, its WSD skills are far from similar to those of speakers: this makes sense, since how speakers disambiguate is still an open question in linguistics.

4. CONCLUSIONS

In this paper we have studied in what ways both disambiguation processes, polysemy and homonymy, differ, demonstrating that context facilitates the correct interpretation of ambiguity; and that homonymy is better processed than polysemy according to the WSD results. Yet, our findings are limited since there were only 44 occurrences with 22 ambiguous units. However, this may be a basis for a deeper reflection on how linguistic information needs to be considered when programming tools such as *Babelfy*.

Some of our recommendations to improve WSD models include the analysis of the types of ambiguity and the different types of relation between meanings (for instance, establishing some rules for sense extension in the case of regular polysemy), to consider linguistic phenomena like collocations and registers used by native speakers during the information-gathering/collecting process (not only linguistic information but also external strategies/para-linguistic?), or to enrich the databases with more heterogenous corpora.

This modest comparative study hopes to be an example on how linguistic knowledge and, more importantly, linguists can contribute to the field of NLP.

ACKNOWLEDGMENTS

This research study was funded by DGA and was supported by the Spanish AEI and Feder (EU) through grant FFI 2017-82460-P. I would also like to thank the anonymous reviewers, as well as my colleague Rosana Villares, whose comments have greatly improved this manuscript.

REFERENCES

Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The Semantic Web. Scientific American. Retrieved from http://www.sciam.com/article.cfm?id=the-semantic-web&page=2

Escandell, M. V. (2008). *Apuntes de semántica léxica*. Madrid: Universidad Nacional de Educación a Distancia (UNED).

Gale, W., Church, K. W. & Yarowsky, D. (1992). Estimating upper and lower bounds on the performance of word-sense disambiguation programs. In H.S. Thompson (Ed.), *Proceedings of the 30th Annual Meeting on Association for Computational Linguistics* (pp. 249-256). Newark: Association of Computational Linguistics.

Haro, J., Ferré, P., Boada, R. & Demestre, J. (2017). Semantic ambiguity norms for 530 Spanish words. *Applied Psycholinguistics*, 38(2), 457-475. doi: 10.1017/s0142716416000266

Hirschberg, J. & Manning, D. (2015) Advances in natural language processing. *Science*, *349*(6245), 261-266. doi: 10.1126/science.aaa8685

Jacobs, A.M. & Ziegler, J.C. (2015). Neurocognitive psychology of visual word recognition. In J.D. Wright (Ed.), *International Encyclopedia of Social and Behavioral Sciences.* (pp. 214-219). Amsterdam: Elsevier.

Jurafsky, D. & Martin, J.H. (2019). *Speech and Language Processing.* (3rd ed. draft). Retrieved from https://web.stanford.edu/~jurafsky/slp3/

López-Cortés, N. (2019). La interpretación subjetiva de la ambigüedad léxica: una aplicación lexicográfica. *LinRed. Lingüística en la Red*, 17, 1-16.

López-Cortés, N. (2021). No todas las ambigüedades son iguales: un estudio sobre la homonimización de la polisemia. In A. Moreno Moreno & M. Torres Martínez (Coords.), *Investigaciones Léxicas. Estados, Temas y Rudimentos* (pp. 302-312). Barcelona: Octaedro.

Moro, A., Raganato, A. & Navigli, R. (2014). Entity linking meets Word Sense Disambiguation: a unified approach. In D. Lin, M. Collins & L. Lee (Eds.), *Transactions of the Association for Computational Linguistics, volume 2* (pp. 231-244). Cambridge, MA: MIT Press.

Morton, H. C. (1994). *The Story of Webster's Third. Philip Gove's Controversial Dictionary and its Critics*. Cambridge: Cambridge University Press.

Navigli, R. (2009). Word Sense Disambiguation: a survey. *ACM Computing Surveys*, 41(2), 1-69. doi: 10.1145/1459352.1459355

Navigli, R. (2015). *BabelNet, Babelfy and beyond: electronic lexicography from machines and the crowd*. Keynote paper presented at *ELex 2015*. Sussex. Video retrieved from https://www.youtube.com/watch?v=mliOk-ZCR1E&t=3067s

Navigli, R. & Ponzetto, S. (2012). BabelNet: The automatic construction, evaluation and application of wide-coverage multilingual semantic network. *Artificial Intelligence*, 193, 217-250. doi: 10.1016/j.artint.2012.07.001

Ouhnana, M. & Kingdom, F.A.A. (2016). Perceptual-binding in a rotating Necker cube: the effect of context motion and position. *Vision Research*, 126, 59-68. doi: 10.1016/j.visres.2016.02.005

Pinker, S. (2007). *The Language Instinct*. New York: Harper Perennial Modern Classics.

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge: MIT Press.

Rodd, J., Gaskell, M. G. & Marslen-Wilson, W.D. (2002). Making sense of semantic ambiguity semantic competition in lexical access. *Journal of Memory and Language*, 46(2), 245-266. doi: 10.1006/jmla.2001.2810

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), 192-233. doi: 10.1037/0096-3445.104.3.192

Schütze, H. (1997). *Ambiguity Resolution in Language Learning. Computational and Cognitive Models* (Doctoral dissertation). Stanford University.

Solé, R., Corominas-Murtra, B., Valverde, S. & Steels, L. (2010). Language Networks: their structure, function and evolution. *Complexity*, *15*(6), 20-26.

Srinivasan, M. & Rabagliati, H. (2015). How concepts and conventions structure the lexicon: cross-linguistic evidence from polysemy. *Lingua*, 157, 124-152. **10.1016/j.lingua.2014.12.004**

Swinney, D.A. (1979). Lexical Access during Sentence Comprehension. (Re)Consideration of Context Effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645-659.

Wasow, T., Perfors, A. & Beaver, D. (2005). The puzzle of ambiguity. In C.O. Orgun & P. Sell (Eds.), *Morphology and the Web of Grammar: Essays in Memory of Steven G. Lapointe* (pp. 265-282). Chicago: CSLI Publications.

Youn, H., Sutton, L., Smith, E., Moore, C., Wilkins, J. F., Maddieson, I., Croft, W. & Bhattacharya, T. (2016). On the universal structure of human lexical semantics. *Proceedings of the National Academy of Sciences*, 113(7), 1766-1771.