

Design of a corpus of stimuli for a psycholinguistic study of lexical ambiguity

Natalia López-Cortés
University of Zaragoza / Spain

Abstract – Lexical ambiguity takes place when a word has more than one meaning. This phenomenon could therefore lead to multiple difficulties in the processing of information; however, speakers deal almost effortlessly with ambiguous units on a daily basis. In order to understand how ambiguous items are processed by speakers, a clear synchronic definition of homonymy and polysemy is needed. In this paper a methodology to gather subjective information about ambiguous words and the relation within their meanings is proposed. Based on this methodology, a corpus of Spanish stimuli is being developed: this corpus consists of words classified as monosemic, homonymous and polysemous via the subjective interpretation of Spanish speakers. This corpus could be used to conduct experimental tasks to determine the behaviour in on-line processing of items with more than one meaning, in order to later design appropriate methods of approaching this complex phenomenon from the point of view of Psycholinguistics.

Keywords – lexical ambiguity; homonymy; polysemy; psycholinguistics; stimuli; synchrony

1. INTRODUCTION¹

The main goal of this paper is to explain the methodology used to develop a corpus of Spanish ambiguous and non-ambiguous words. This corpus is thought to be the basis for experimental approaches to the processing and storage of meanings in long-term memory: that is to say, the words in this corpus could be used as material for psycholinguistic research. This paper therefore gives an account of the study of lexical ambiguity from a psycholinguistic point of view, placing emphasis on the importance of having corpora of materials and stimuli that have been classified taking into account subjective interpretation of words and its meanings.

Firstly, the theoretical framework is presented in Section 2, focusing on the definition of lexical ambiguity, polysemy and homonymy, and on the already existing

¹ This research 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 whose comments have greatly improved this manuscript.



psycholinguistic analysis of words with more than one meaning. This will show that there is a gap to fill in the study of lexical ambiguity in Spanish, since material based on subjective classification of homonymy and polysemy is needed. Secondly, the methodology used to design the corpus is explained in Section 3. The biggest contribution of this corpus is the subjective classification of ambiguous words as homonymous or polysemous, as well as the classification of words as non-ambiguous. To gather this information, questionnaires were used. Variable control of these lexical units will also be presented in Section 3.2. Then, a description of the current corpus and its possible applications is provided, as well as a brief comparison of the corpus with the definition of homonymy from a lexicographic point of view, in order to prove the importance of subjective measures. Finally, some future lines of research that could be undertaken to study lexical ambiguity in Spanish are sketched in Section 5.

2. THEORETICAL FRAMEWORK

2.1. *Lexical ambiguity: Different definitions of a semantic phenomenon*

Lexical ambiguity is a linguistic phenomenon that has been broadly studied. It takes place when a single lexeme has two or more meanings, as it occurs in Spanish with *llama* ‘flame’ and *llama* ‘llama’ and in English with *rabbit*-ANIMAL and *rabbit*-MEAT. It is therefore opposed to the concept of monosemy, in which a lexical form is mapped only to one meaning.

From a diachronic point of view (i.e. when the etymological origin and historical evolution of the words are considered), two types of lexical ambiguity are usually established in the literature: homonymy and polysemy. Homonymy takes places when two different words happen to converge in a single linguistic form (e.g. the Latin word *flamma* and the Quechuan word *llama*, which converge in the Spanish word *llama*). A polysemous word is produced when a word extends its meaning to designate new realities or entities (e.g. *pluma* ‘feather’ and *pluma* ‘pen’ in Spanish). Gutiérrez Ordóñez (1989: 125) claims that homonymous words can be described in terms of a phonetic convergence, whilst polysemous items, as a result of a semantic diversification.

The different types of ambiguity are reflected in lexicography. The differences in the diachronic evolution of words are depicted in the dictionaries in two ways:

homonymous words are presented under different, separated lexical entries, whereas polysemous units are presented in a single entry, where their multiple meanings are listed.

Although all the diachronic data about lexical ambiguity is interesting, it is not pertinent when studying the processing and storage of lexical units from a psycholinguistic approach. The etymological origin of words does not correlate with the psychological interpretation of ambiguous words by speakers (López-Cortés 2019).² In other words, the historical and etymological evolution of a word does not have a psychological correlate: speakers do not need to know the etymological origin of words; as a matter of fact, they are normally ignorant of it.

It is therefore important to consider the psychological interpretation of ambiguous words when studying the phenomenon of lexical ambiguity from a psycholinguistic point of view. The reason for this is that when speakers process a word, the information they access is the one stored in their memory, and the nature of that information is subjective. According to this synchronic approach, homonymy takes place when a word possesses more than one meaning and those meanings are not related in any way. By contrast, polysemy occurs when a word implies more than one meaning but those meanings are in some way related to one another. Rodd *et al.* (2002) rightly suggest that homonymous words have different meanings and polysemous words, different senses.

2.2. *Psycholinguistic approach to ambiguous units*

The lexical ambiguity phenomenon has played a key role in psycholinguistic research over the last decades. The fact that a single lexical form can transmit a variety of meanings, related or unrelated to one another, arouses interest mainly when trying to understand how speakers process words. Several studies investigating this topic have been carried out. The most common task to study the processing of lexical units is lexical decision tasks in which the participant needs to decide if the stimulus shown on the screen is a real word of their native language or a non-word (i.e. a string of letters that do not correspond to an actual word). When conducting this type of task, some authors discovered lower reaction times when processing an ambiguous stimulus (see, among others, Millins and Button 1989; Hino and Lupker 1996; Hino *et al.* 2002; Lin and Ahrens

² López-Cortés's (2019) only found processing phenomena such as those presented in Section 2.2 when the stimuli were classified following subjective metrics.

2010). In the last decade, this ambiguity advantage has been revised and a different behaviour for homonymy and polysemy has been identified. It has been discovered that the polysemous items were the only ones generating lower reaction times in lexical decision tasks (Rodd *et al.* 2002; Beretta *et al.* 2005; Klepousniotou and Baum 2007). Thus, the ambiguity advantage was reformulated into the polysemy advantage and the homonymy disadvantage.³

These differences in processing are interesting, especially since they are thought to point to differences in the way words are stored in the mental lexicon. If differences between homonymy and polysemy are found in lexical decision tasks, then these lexical units are being accessed differently in the mental lexicon. Many approaches to the storage of lexical ambiguity have been suggested (see Falkum and Vicente 2015 for a review). Nonetheless, the most extended model opts for a representation in separated, autonomous entries for homonymous words and a representation in a single entry for polysemous words.⁴

This model is consistent with the data about the processing of ambiguous words obtained in the lexical decision tasks. The unrelated meanings of homonymy are stored in separated entries of the mental lexicon and a competition for activation between them takes place during lexical access. As a result, higher reaction times are generated, and the homonymy disadvantage is explained. In contrast, when recognising a polysemous word, a single entry is accessed and consequently there is no competition for activation. This entry is richer and more complex than the one for homonymous items, since it should contain some sort of basic meaning that could be extended to express the specific senses of the word. The issue of the representation of polysemy in the mental lexicon has been a much-disputed subject within the field of psycholinguistics and there is still

³ It is important to note that the data that proves a differential behaviour for homonymy and polysemy is usually based on English stimuli. Although these phenomena have been replicated in other languages (see, for example, Lin and Ahrens 2010), when conducting an experiment in Spanish the results are not clear. Haro *et al.* (2017a) were not able to find a difference between homonymy and polysemy in their experimental tasks. It can be therefore claimed that the issue of the processing of ambiguous units is still controversial and needs further reflection, especially if the comparison between languages is considered. Furthermore, the effects could change not only depending on the language used but also on the type of experimental task selected (Eddington and Tokowicz 2015).

⁴ One of the most interesting things to consider when analysing this data is that the distinction between homonymy and polysemy may not be so strict; it is more likely to be somehow more gradual and less discrete. When conducting an experimental task, it is essential to determine the classification of the items and researchers need to establish criteria to do so, but since this may not be the most ecological solution the data needs to be examined critically. Here we propose one of these criteria to classify a subjective phenomenon as lexical ambiguity in an objective way (see Section 3.1).

disagreement about how this single entry is structured. The most extended approach is the *core meaning* theory (Klepousniotou and Baum 2007), although it has been strongly challenged in recent years by some authors (Foraker and Murphy 2012).

2.3. *Why is a corpus needed?*

All in all, the type of ambiguity of a word (i.e. whether it is polysemous or homonymous) affects its processing and storage. However, as already mentioned in Section 2.1, the type of ambiguity can be measured from diachrony (the origins of a lexical unit) or from synchrony (the interpretation of the relation between its meanings). Both criteria are not always equivalent. For instance, the word *catarata* (which can mean either ‘waterfall’ or ‘cataract’) is polysemous in Spanish since it has got one single Latin origin (*cataracta*). However, from a synchronic point of view, its meanings are interpreted as unrelated for which it can be considered homonymous.⁵ It therefore follows that it is essential to determine which approach is needed to study the processing and the storage of ambiguous units.

We believe that subjective information is what is relevant when studying a semantic phenomenon from a psycholinguistic point of view. As it has already been mentioned in Section 2.1, speakers are normally not aware of etymology, and therefore of diachrony. As a consequence, in order to conduct experiments, subjectively-classified stimuli are needed, since that subjective information which is stored in the lexicon is what speakers need to have access to in order to communicate. Creating a corpus of these characteristics to study the behaviour of homonymy and polysemy is the main goal of the present, ongoing research.

It is important to point out that there are already some Spanish subjective corpora published. Estévez (1991) collected 214 subjectively-classified ambiguous words, which were then classified as homonymous and polysemous following the lexicographic criteria. Domínguez *et al.* (2001) focused entirely on polysemy, proposing 100 polysemous words. Gómez-Veiga *et al.* (2010) gathered information about 113 ambiguous words and different variables, such as frequency or dominance of meanings,

⁵ The methodology used to gather this subjective interpretation is explained in Section 3.1.

but there was no further classification of those items considering the relation between their meanings.

The authors of all these corpora were aware of the fact that subjective metrics are the ones to consider when studying ambiguity from a psycholinguistic point of view. For instance, Domínguez *et al.* (2001: 65) claim that, although the dictionary directly offers the number of meanings (*acepciones*), that number is not psychologically relevant. However, these authors, as well as Estévez (1991), use the dictionary to determine whether a word is homonymous or polysemous and only consider subjective interpretation with regard to the number of meanings (in other words, to determine if an item is ambiguous or not). Besides, these materials lack a set of non-ambiguous words with which the ambiguous words can be compared, as already noted by Haro *et al.* (2017b).

The most recent efforts to design an ambiguity corpus are the ones by Fraga *et al.* (2017) and Haro *et al.* (2017b). The *Spanish Ambiguous Words Database* (SAW) by Fraga *et al.* (2017) is an interesting approach to the definition of ambiguity, since it seems to prove, via a meaning retrieval task, that the information contained in the dictionaries is quite similar to the meanings that speaker have stored in their lexicon. The participants of this study had to write meanings of different ambiguous words and those meanings were then compared, through a Pearson correlation, with the information in the lexicographic entries of the most common Spanish dictionary (*Diccionario de la Lengua Española*). The originality of this work is undeniable and its implications can be widely discussed.⁶ However, the only metric taken into account to classify words as polysemous and homonymous is, once again, the lexicographic criterion, which means that the items are classified according to their etymological origin.

Haro *et al.*'s corpus (2017b) consists of 530 words. The most interesting contribution of this work is the fact that a methodology to identify homonymy and polysemy from a subjective point of view is proposed. Haro *et al.* (2017b) present two different subjective variables: NOM (number of meanings) and ROM (relatedness of meanings). The latter variable is obtained through a Likert scale: participants were asked

⁶ There are some works that seem to prove the opposed view: the dictionary approach and the subjective, synchronic classification do not correlate (Haro *et al.* 2015; López-Cortés 2019). Some information regarding this topic is presented in Section 4.1. It is important to know that the objective of the aforementioned research and the one made by Fraga *et al.* (2017) are not equivalent, which could explain the contradictory results.

whether the meanings of a word were related and they had to select a value from 1 to 9. This is an effective way to determine the type of ambiguity of an ambiguous word and, most importantly, it is based on the interpretation of speakers. The methodology and the data analysis used by these authors are different from those presented here. A combination of both approaches could be ideal to expand the corpus and to gather more experimental stimuli in Spanish.⁷

3. CORPUS DESIGN

3.1. Word classification

The most important part of the corpus design was the subjective classification of words, as ambiguous-monosemic and as homonymous-polysemous. Such classification was obtained by means of questionnaires which allowed us to gather subjective and synchronic data of words, which would later on be used as stimuli for psycholinguistic experiments. It must be noted that the methodology was consistent throughout the corpus design: the same type of questionnaire was used and the data were analysed following always similar criteria.

The stimuli corpus was designed based on data obtained through 21 questionnaires, filled between 2015 and 2019 by a total number of 716 native Spanish speakers who gave their explicit consent to participate in the experimental session. Each questionnaire had an average response of 34.09 answers with a standard deviation (SD) of 16.36. The current corpus, which is still being developed, has information about 336 Spanish words.

The questionnaires were designed using *GoogleForms* and consisted of 15–20 words each. The structure of this questionnaire is displayed in Figure 1.

⁷ It is also important to note that all existent corpora are the result of psychological investigations and are therefore made by researchers working on this discipline. Nonetheless, it could be useful to have a linguistic basis to adequately design or interpret data related to a semantic phenomenon such as lexical ambiguity. For this reason, a corpus like the one presented here could be a good complement to previous works.

Los significados de las palabras

*Obligatorio

BARRA

¿Crees que esta palabra tiene un significado o varios? *

☐ Un significado

☐ Varios

En caso de que hayas respondido "varios", ¿crees que los significados más habituales de esta palabra están relacionados entre sí?

☐ Sí, los significados están relacionados

☐ No, son significados muy distintos

ATRÁS SIGUIENTE

Página 2 de 21

Figure 1: Screenshot of one of the questionnaires used to develop the corpus illustrating the Spanish word *barra* meaning ‘stick’ or ‘counter’ (among other meanings)

A word is presented, followed by two questions:

- (i) *Do you believe this word has one meaning or more than one meaning?*
- (ii) *In case you answered “more than one meaning,” do you believe the most common meanings of this word are related?* The possible answers to this question were *Yes, meanings are related* and *No, the meanings are very different*.

With these questionnaires two values were obtained: whether the word is monosemic or ambiguous (question 1) and whether the meanings of the words are considered to be related (polysemy) or not (homonymy) (question 2).⁸

The first words selected to start the corpus were taken from Gómez-Veiga *et al.* (2010). As it has been shown, these researchers did not consider the differentiation between homonym and polysemy and thus their words needed further classification. Then, some words from Haro *et al.* (2017a) and experimental material from Cueto *et al.* (1997) were also employed. However, these corpora were used as a source for material for word selection, but those items were always classified using our own methodology. In this way, total coherence in the design of the corpus was assured. Later on, as the corpus was being designed, new words were added by different means: experimental design and unexpected interpretation made by participants. All the new words were always classified through the questionnaires.

⁸ One of our future lines of research is to perform meaning retrieval tasks (as in Fraga *et al.* 2017) in order to collect the most frequent meanings of these ambiguous units. However, at this stage of corpus design, our main goal was to determine a methodology to express, as objectively as possible, the opposition between monosemy-ambiguity and homonymy-polysemy.

The basis of the analysis procedure was to apply the same objective criteria to all subjective data obtained in the questionnaires. For a word to be included in one category (monosemy, polysemy or homonymy) a minimum agreement of 60% in the answers of all participants had to be reached. Some examples of this classification are presented next in (1) to (4).

- (1) *avestruz* ('ostrich'): monosemic with 80% of agreement
- (2) *flamenco* ('flamingo'-'flamenco'): ambiguous with 88% of agreement
- (3) *estrella* ('star'-'famous person'): polysemous with 75% of agreement
- (4) *banco* ('bank'-'bench'): homonymous with 95.3% of agreement

What is interesting about having this percentual information is that it reflects the fact that ambiguity seems to be a scale: the relation between meanings is gradual (see fn 4). Interpreting ambiguity this way also shows how subjective interpretation of lexical units defines the semantic phenomenon: the semantic information stored depends on the individual speakers, since not all of them interpret the items in a similar way. All this information can be useful when analysing experimental data.

Establishing a minimum percentage of agreement also helps eliminate those words that cannot be classified since their values do not reach the minimum percentage established, as it occurs with (5) and (6).

- (5) *carta* ('card'-'letter'-'menu'): between monosemy (44%) and ambiguity (56%)
- (6) *grano* ('grain'-'spot'): between polysemy (54.5%) and homonymy (45.5%)

These words have not been included in the corpus yet and further research is needed to properly classify them.⁹ However, they lend further support to the idea that ambiguity is a gradual phenomenon and that it depends on the interpretation of each speaker.

3.2. *Participants and procedure*

A total number of 716 Spanish native speakers took part in the filling in of the questionnaires. Since the tool *GoogleForms* allows to collect information from the questionnaires online, some participants filled in the questionnaires from their homes but most of them did it on-site. The most common profile of participant corresponded to students of the degrees of Spanish Philology and Classical Studies at the University of

⁹ We believe that the most adequate approach is to study those words from a linguistic point of view: analysing the semantic features of these units may clarify why these words are hard to classify.

Zaragoza. The age range was between 18 and 25 and all of them lived in the province of Zaragoza (Aragón, Spain).

Participants were told that their answers would be used statistically and they were asked to answer according to their own interpretation as Spanish speakers. They were aware that there was no time control and that they could use as much time as they needed to fill in the questionnaire. The fact that there were no right or wrong answers was specially stressed, so that they would answer according to their own interpretation.

The questionnaire was normally presented to the participants after they had already completed another task. A soundproofed room was used and the questionnaire was filled in via a laptop with Internet connection. The duration of this session varied depending on individual speed, but it was never longer than 15 minutes.

3.3. *Variable control*

The effect of ambiguity has been studied for decades now and some authors have explored the possibility that there are some variables that could interact or even interfere with the processing of words with more than one meaning. Different tasks and experiments have been carried out in the last years by researchers to determine which these variables are.

There are different approaches, methodologies and points of view but the variables most commonly studied in relation with ambiguity are the ones that follow: frequency (Rubenstein *et al.* 1970; Gernsbacher 1984; Cuetos *et al.* 1997 and more recently Jager *et al.* 2016, among others), familiarity (Gernsbacher 1984), imaginability (Cuetos *et al.* 1997) and concreteness (Tokowicz and Kroll 2007; Jager and Cleland 2016). Out of all these variables, frequency is, by far, the most amply studied. Nonetheless, its effect on ambiguity is not clear: in most cases the influence of frequency interacts with the type of task or the experimental design. For this reason, the most common approach is to control this variable when conducting an experiment: that is, using items with similar frequency to make sure that the frequency is not accountable for any processing effects that arise.

The objective when designing the corpus presented in this paper was to control for all these variables, in order to have information only about the number of meanings and their relationship. In this way it can be guaranteed that if an effect is found in an experimental task it will be caused by the ambiguity values and not by other lexical or subjective variables.

The data for all these variables was extracted from different already existent corpora: relative frequency and absolute frequency¹⁰ from the NIM corpus (Guasch *et al.* 2013) and familiarity, imaginability and concreteness from the EsPal corpus (Duchon *et al.* 2013). The information related to each word was included in the corpus and was later analysed in three different groups: (i) homonymy-monosemy, (ii) polysemy-monosemy, and (iii) polysemy-homonymy. This analysis was to check that there were no statistically significant differences regarding the variables that could affect the ambiguity effect.

The non-parametric Wilcoxon test was conducted to compare the variables. The level of significance (p) was established at 0.05. However, statistically significant results were not obtained in either of the groups, as shown in Table 1, where the result of the Wilcoxon test is presented in the first column (V -stat.).¹¹ It can be therefore claimed that there are not statistically significant differences between groups regarding these variables, which means that none of them should have an effect in experimental tasks.¹²

	Homonymy-Monosemy		Polysemy-Monosemy		Polysemy-Homonymy	
	<i>V-stat.</i>	<i>p-value</i>	<i>V-stat.</i>	<i>p-value</i>	<i>V-stat.</i>	<i>p-value</i>
<i>Relative frequency</i>	1,810	0.5394	1,655	0.2085	1,824	0.5786
<i>Absolute frequency</i>	1,811	0.5421	1,655	0.2081	1,822.50	0.5743
<i>Familiarity</i>	2,171.50	0.1957	1,779	0.9804	2,218	0.1351
<i>Imaginability</i>	2,021	0.2936	1,679	0.5167	1,897	0.7624
<i>Concreteness</i>	1,558	0.1791	2,082	0.2657	1,658	0.3613

Table 1: Results of the variable control

4. CORPUS DESCRIPTION

The corpus currently consists of 336 words, subjectively classified into three groups: monosemy, homonymy and polysemy. It is therefore divided in three sections: monosemic stimuli (88 words), homonymous stimuli (88 words) and polysemous stimuli (160 words).

¹⁰ Relative frequency is the appearance of the word in parts per million whereas absolute frequency is the total number of appearances of the word in the corpus, as explained by Guasch *et al.* (2013).

¹¹ One anonymous reviewer suggests including a comparison between ambiguity and monosemy in Table 1. However, we do not have this data at the moment, since we are mainly interested in the processing of homonymy and polysemy. This will be, however, considered in future research.

¹² One of the most important steps when designing experimental tasks is controlling variables that can have an effect on the results. For this reason, these variables should always be controlled for before carrying out any tasks. The variable control presented here works for our research since all these items were used in lexical decision tasks and, as an essential part of the corpus design, we thought it interesting to show this process in the present paper. The data presented here can work as a basis, but it is highly recommendable to repeat the controlling process depending on each researcher's experimental design and objectives.

The words in each section are ordered by agreement degree (from higher to lower) and different variables for each word are then listed: frequency (obtained from Guasch *et al.* 2013), familiarity, imaginability and concreteness (obtained from Duchon *et al.* 2013). These latter variables were measured by researchers through a Likert scale, where participants had to decide how familiar, imageable or concrete a word was in a scale from 0 to 7.

The most interesting additions to this corpus are the following: firstly, the incorporation of the homonymy-polysemy classification, based on a subjective interpretation obtained through questionnaires. This data is reflected with a percentage of the agreement in the classification, which allows us to assess whether there are differential effects of processing for words that fall within the same category but have classifications that vary greatly in agreement.

Secondly, the fact that the information about reaction times is added to the corpus is also interesting. Each word is followed by the mean of the reaction times that the item produces in lexical decision tasks. This measurement is presented in milliseconds and was obtained by conducting a series of lexical decision task with the material of the corpus.¹³

This information (classification, agreement and reaction times) is the major contribution of this corpus. In Table 2, a summary of the all data is presented.

	<i>Homonymy</i>		<i>Polysemy</i>		<i>Monosemy</i>	
	Mean	SD	Mean	SD	Mean	SD
<i>Reaction time</i>	763.1	95.75	737.17	78.11	730.14	68.46
<i>Relative frequency</i>	57.28	99.35	85.09	154.66	63.68	104.05
<i>Logarithm</i>	1.36	0.58	1.51	0.59	1.37	0.63
<i>Absolute frequency</i>	322.43	559.28	4790	870.63	358.49	585.72
<i>Familiarity</i>	5.27	1.62	5	1.93	4.86	2.03
<i>Imaginability</i>	4.54	1.91	4.51	2.03	4.27	2
<i>Concreteness</i>	4.08	1.42	4.27	1.67	4.29	1.91

Table 2: Summary of the data presented in the corpus, with the mean of the values and their standard deviation (SD)

With all the data gathered, we believe it is relevant to point out once again how the information compiled in the dictionary is not the same as the one in the long-term memory of native speakers, at least regarding the classification of ambiguous words, such as

¹³ This data consists of only a small sample, but it is a preliminary approach to investigating the way in which these units are processed. Other researchers, such as González-Nosti *et al.* (2014), had already collected this measure as valuable information in previous corpora.

homonymy or polysemy (as already shown by Haro *et al.* 2015 and López-Cortés 2019; but somehow contrary to Fraga *et al.* 2017).¹⁴

Homonymy has been considered to be a far less frequent phenomenon than polysemy, since it is hard for two non-related words to converge in form. However, our psychological data reveals that homonymy seems to be more common than expected when it is measured from a diachronic point of view.

In (7), the homonymous items from our corpus are presented. These words are considered by our participants to have multiple non-related meanings. The units that are also classified as homonymous in the *Diccionario de la Lengua Española* are in bold.

- (7) apéndice, artículo, banco, **banda**, **bolsa**, **borde**, **bota**, **bote**, **cabo**, **cala**, cámara, campaña, **canto**, caña, **cardenal**, carrera, caso, catarata, **celo**, chisme, **chorizo**, chuleta, clase, **coco**, **cola**, **cólera**, **colonia**, **coma**, compañía, concierto, cuadrado, **cubo**, cura, **duelo**, esposa, estación, estado, ficha, flamenco, general, genio, **golfo**, **grado**, grano, gravedad, guion, hábito, **heroína**, **jota**, ladrón, lata, **lima**, línea, **lira**, lista, **mango**, marca, marea, **media**, medio, mina, monitor, mono, **muelle**, muñeca, nota, obra, palma, papel, parábola, parte, partida, partido, pasta, pendiente, **pez**, piña, pluma, **pompa**, puesto, pupila, rana, segundo, servicio, taco, **tapa**, **tela**, tienda.

As can be seen, there are more subjectively-classified items as not having related meanings (i.e. homonymous) than would be expected if the dictionary approach was taken exclusively: out of 88 homonymous words from the present corpus only 31 are also homonymous based on their etymology, while the other 57 are reflected in the dictionary as polysemous. In other words, 57 words that are etymologically polysemous have been reinterpreted as homonymous through our questionnaires. That means that, when taking a subjective approach, homonymy increases and speakers tend to interpret words as having unrelated meanings more frequently than expected.

The data show that a corpus design that takes into account the psychological, subjective differences between types of ambiguity is indeed a useful tool.

¹⁴ It is important to reiterate that the objective of Fraga *et al.*'s work (2017) was to check whether the meanings speakers retrieve from memory are the same as the ones that dictionaries reflect. Their results show that there is a positive correlation between these two measures. However, the differentiation between homonymy and polysemy was not taken into account, at least not in terms of relation between meanings, since these authors considered only the number of meanings and the semantic information that each lexical form gathers.

5. APPLICATIONS AND FUTURE LINES OF RESEARCH

This corpus is a starting point to investigate lexical ambiguity in Spanish from a psycholinguistic point of view. Once finished, it can be used to develop experimental tasks in Spanish, as it is a source for material that has been carefully controlled. Moreover, since it has been proved that ambiguity is not a homogenous phenomenon (Klepousniotou and Baum 2007), the classification of homonymy and polysemy based on subjective interpretation can be key to a robust experimental design. We believe that the most important contribution of our research is the reflection of how the different types of ambiguity could be approached through an objective measurement of subjective interpretation which allows us to obtain a scale of values.

Having a corpus based on Spanish stimuli can be key to establish whether the processing phenomena found in English (the advantage of polysemy and the disadvantage of homonymy) are also produced in other languages.

This corpus can also be the basis of a linguistic study of words with more than one meaning. One of our lines of research is to determine the nature of the relation between meanings by studying the features that characterise polysemy and homonymy. To do so, a meaning retrieval task should be carried out (see fn 10).

Further work needs to be done to expand the corpus: more ambiguous and monosemic nouns should be subjectively classified in order to design new experimental tasks that allow us to understand the processing of different meanings. It would also be interesting to start gathering new categories such as verbs or adjectives, or even items which show ambiguity within their category (as it occurs with the Spanish word *pobre* which can either be interpreted as a noun ‘a poor man’ or an adjective ‘poor’).

We believe that designing experimental material based on subjective approaches, that is, taking into account the interpretation of speakers, is the proper way to move forward if we want to fully understand the nature of the processing mechanisms related to lexical ambiguity in particular and lexical units in general.

REFERENCES

- Beretta, Alan, Robert Fiorentino and David Poeppel. 2005. The effects of homonymy and polysemy on lexical access: An MEG study. *Cognitive Brain Research* 24/1: 57–65.
- Cuetos, Fernando, Alberto Domínguez and Manuel de Vega. 1997. El efecto polisemia: Ahora lo ves otra vez. *Cognitiva* 9/2: 175–194.
- Domínguez, Alberto, Fernando Cuetos and Manuel de Vega. 2001. 100 palabras polisémicas con sus acepciones. *Revista Electrónica de Metodología Aplicada* 6/2: 63–84.
- Duchon, Andrew, Manuel Perea, Nuria Sebastián-Gallés, Antonia Martí and Manuel Carreiras. 2013. EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods* 45/4: 1246–1258.
- Eddington, Chelsea and Natasha Tokowicz. 2015. How meaning similarity influences ambiguous word processing: The current state of the literature. *Psychonomic Bulletin & Review* 22/1: 13–37.
- Estévez, Adelina. 1991. Estudio normativo sobre ambigüedad en castellano. *Cognitiva* 3/2: 237–271.
- Falkum, Ingrid Lossius and Agustín Vicente. 2015. Polysemy: Current perspectives and approaches. *Lingua* 157: 1–16.
- Foraker, Stephani and Gregory Murphy. 2012. Polysemy in sentence comprehension: Effects of meaning dominance. *Journal of Memory and Language* 67/4: 407–425.
- Fraga, Isabel, Isabel Padrón, Manuel Perea and Montserrat Comesaña. 2017. I saw this somewhere else. The Spanish Ambiguous Words (SAW) database. *Lingua* 185: 1–10.
- Gernsbacher, Morton Ann. 1984. Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General* 113/2: 256–281.
- Gómez-Veiga, Isabel, Nuria Carriedo, Mercedes Rucián and Óscar J. Vila. 2010. Estudio normativo de ambigüedad léxica en castellano en niños y en adultos. *Psicológica* 31: 25–47.
- González-Nosti, María, Analía Barbón, Javier Rodríguez-Ferreiro and Fernando Cuetos. 2014. Effects of the psycholinguistic variables on the lexical decision task in Spanish: A study with 2765 words. *Behaviour Research Methods* 46/2: 517–525.
- Guasch, Marc, Roger Boada, Pilar Ferré and Rosa Sánchez-Casas. 2013. NIM: A Web-based Swiss Army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods* 44/3: 756–771.
- Gutiérrez Ordóñez, Salvador. 1989. *Introducción a la Semántica Funcional*. Madrid: Síntesis.
- Haro, Juan, Pilar Ferré, Roger Boada and Josep Demestre. 2015. Ambiguity advantage depends on how ambiguous words are categorized. Poster presented at the *XII International Symposium of Psycholinguistics*, University of València, 3rd July, 2015.
- Haro, Juan, Josep Demestre, Roger Boada and Pilar Ferré. 2017a. ERP and behavioral effects of semantic ambiguity in a lexical decision task. *Journal of Neurolinguistics* 44: 190–202.
- Haro, Juan, Pilar Ferré, Roger Boada and Josep Demestre. 2017b. Semantic ambiguity norms for 530 Spanish words. *Applied Psycholinguistics* 38/2: 457–475.

- Hino, Yasushi and Stephen Lupker. 1996. Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance* 22/6: 1331–1356.
- Hino, Yasushi, Stephen Lupker and Penny Pexman. 2002. Ambiguity and synonymy effects in lexical decision, naming and semantic categorization tasks: Interactions between orthography, phonology and semantics. *Journal of Experimental Psychology: Learning Memory and Cognition* 28/4: 686–713.
- Jager, Bernardet and Alexandra Cleland. 2016. Polysemy advantage with abstract but not concrete words. *Journal of Psycholinguistic Research* 45/1: 143–156.
- Jager, Bernardet, Matt Green and Alexandra Cleland. 2016. Polysemy in the mental lexicon: Relatedness and frequency affect representational overlap. *Language Cognition and Neuroscience* 31/3: 425–429.
- Klepousniotou, Ekaterini and Shari R. Baum. 2007. Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics* 20/1: 1–24.
- Lin, Chien-Jer Charles and Kathleen Ahrens. 2010. Ambiguity advantage revisited: Two meanings are better than one when accessing Chinese nouns. *Journal of Psycholinguistics Research* 39: 1–19.
- López-Cortés, Natalia. 2019. La interpretación subjetiva de la ambigüedad léxica: Una aplicación lexicográfica. *LinRed: Lingüística en la Red* 17: 1–16.
- Millis, Michelle L. and Scoti B. Button. 1989. The effect of polysemy on lexical decision time: Now you see it, now you don't. *Memory and Cognition* 17/2: 141–147.
- Real Academia Española. 2018. *Diccionario de la Lengua Española*. <http://www.rae.es/rae.html>
- Rodd, Jennifer, Gareth Gaskell and William Marslen-Wilson. 2002. Making sense of semantic ambiguity semantic competition in lexical access. *Journal of Memory and Language* 46/2: 245–266.
- Rubenstein, Herbert, Lonnie Garfield and Jane A. Millikan. 1970. Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior* 9/5: 487–494.
- Tokowicz, Natasha and Judith F. Kroll. 2007. Number of meanings and concreteness: Consequences of ambiguity within and across language. *Language and Cognitive Processes* 22/5: 727–779.

Corresponding author

Natalia López-Cortés
 Departamento de Lingüística General e Hispánica
 Facultad de Filosofía y Letras
 Universidad de Zaragoza
 C/ San Juan Bosco 7
 50009, Zaragoza (Spain)
 e-mail: natlop@unizar.es

received: October 2019
 accepted: February 2020