TITLE:

Semantic ambiguity norms for 530 Spanish words

AUTHORS:

Juan Haro, Pilar Ferré, Roger Boada, Josep Demestre

Research Center for Behavior Assessment and Department of Psychology; Universitat

Rovira i Virgili. Tarragona. Spain

RUNNING HEAD:

Semantic ambiguity norms

CORRESPONDING AUTHOR:

Juan Haro

Department of Psychology and CRAMC. University Rovira i Virgili

Crta. De Valls s/n., 43007-Tarragona. Spain

e-mail: juan.haro@urv.cat

Telephone Number: 34-977-558567

Abstract

This study presents semantic ambiguity norms for 530 Spanish words. Two subjective measures of semantic ambiguity and two subjective measures of relatedness of ambiguous words' meanings (ROM) were collected. In addition, two objective measures of semantic ambiguity were included. Furthermore, subjective ratings were obtained for some relevant lexico-semantic variables, such as concreteness, familiarity, emotional valence, arousal and age-of-acquisition. In sum, the database overcomes some of the limitations of the published databases of Spanish ambiguous words; in particular, the scarcity of measures of ambiguity, the lack of ROM measures, and the absence of a set of unambiguous words. Thus, it will be very helpful for researchers interested in exploring semantic ambiguity as well as for those using semantic ambiguous words to study language processing in clinical populations.

Keywords

Language is full of words with more than one meaning (e.g., the word *bat*, which means *nocturnal flying mammal* as well as *wooden club used in baseball*), the so-called ambiguous words. Indeed, it is estimated that approximately 44% of English words are ambiguous, a quantity that increases considerably for high frequency words (Britton, 1978). Given their abundance, understanding how ambiguous words are represented and processed by the human mind is necessary for a complete explanation of human language processing and representation.

A unique property of ambiguous words is that they show a particular one-to-many mapping between their spelling and their meanings. This property makes them very interesting and useful for psycholinguistic research, especially for studying how words are processed and represented. For example, research on semantic ambiguity has contributed importantly to elucidate how semantics and orthography interact during word recognition (e.g., Balota, Ferraro, & Connor, 1991; Hino, Lupker, & Pexman, 2002) or how a previous context affects meaning activation during reading (e.g., Swinney, 1991; Van Petten & Kutas, 1987). In addition, ambiguous words have also been employed in other fields of research. Just to name a few, they have been used to investigate implicit memory and false memories (Eich, 1984; Hutchison & Balota, 2005), depression (Hertel & El-Messidi, 2006), personality disorders (Mathews, Richards, & Eysenck, 1989) or autism (Hala, Pexman, & Glenwright, 2007).

Focusing on word recognition research, a common finding is that ambiguous words are recognized faster than unambiguous words in the lexical decision task (LDT) (e.g., Hino & Lupker, 1996; Jastrzembski, 1981; Kellas, Ferraro, & Simpson, 1988; Millis & Button, 1989). This processing advantage for ambiguous words in the LDT (hereafter, *ambiguity advantage*) has been a challenge for most of the models of word recognition, especially for classical models and parallel distributed processing (PDP) models. On the one hand, classical

models were unable to provide an explanation for this effect, as they assumed that the information related to the meaning(s) of a word (i.e., semantic information) should not affect word recognition. Thus, to account for the ambiguity advantage, some changes in these models were proposed (e.g., logogen models, Jastrzembski, 1981; and serial-search models, Rubenstein, Garfield, & Millikan, 1970). For instance, it was suggested that ambiguous words might be represented by more than one lexical entry, each one connected to a distinct meaning of the word. Consequently, it would be more probable to quickly select one of the lexical entries of an ambiguous word than to select the single entry of an unambiguous word, resulting in a processing advantage for ambiguous words.

On the other hand, PDP models were neither able to account for the ambiguity advantage. In fact, PDP models predicted a processing disadvantage for ambiguous words, given the assumption that ambiguous words have an inconsistent one-to-many mapping between their orthographic representation (e.g., *bark*) and their semantic representations (e.g., *the sharp cry of a dog* and *the outside covering of a tree*). This inconsistent mapping would produce a slower semantic coding for ambiguous words, probably due to a process of competition between their multiple semantic representations (e.g., Azuma & Van Orden, 1997). Thus, to provide an explanation for the ambiguity advantage consistent with PDP models, some authors have proposed that recognition times in the LDT do not reflect semantic coding, but orthographic coding (e.g., Hino & Lupker, 1996). Accordingly, since ambiguous words have multiple semantic representations, they would benefit from a large semantic feedback from the different meaning representations to their orthographic representation during word recognition, speeding up the orthographic coding and so then recognition and response times.

Although the above-mentioned accounts of the ambiguity advantage assume that this phenomenon is a result of words having multiple meanings, some studies conducted during

the last decade strongly challenged this assumption (e.g., Armstrong & Plaut, 2011; Klepousniotou & Baum, 2007; Rodd, Gaskell, & Marslen-Wilson, 2002). Such studies examined words with multiple related meanings (i.e., polysemes; e.g., *newspaper*), words with multiple unrelated meanings (i.e., homonyms; e.g., *bat*), and unambiguous words by using the LDT. They reported a disadvantage for homonyms in comparison to unambiguous words, and an advantage for polysemes with respect to unambiguous words and to homonyms. Thus, these studies showed that the degree of relatedness between the different meanings of a word, rather than the number of meanings, facilitates word recognition. To account for this evidence, Rodd, Gaskell, and Marslen-Wilson (2004) suggested that the multiple semantic representations of a word with unrelated meanings would compete during word processing, slowing down their recognition in the LDT; in contrast, the related semantic representations of a word with related meanings would facilitate the recognition of the word. However, it should be noted that some findings are at odds with this relatedness-of-meanings advantage. Indeed, several studies have found an advantage of the same magnitude for polysemes and homonyms in comparison to unambiguous words (e.g., Hino, Kusunose, & Lupker, 2010; Hino, Pexman, & Lupker, 2006).

In view of these conflicting experimental findings, more research is needed for a complete understanding of how ambiguous words are processed and represented, as well as to identify the variables that influence the processes and mechanisms involved in their recognition. However, before conducting a study on semantic ambiguity, researchers have to face several important issues. Three of them are, in our view, particularly critical: 1) selecting a proper set of ambiguous words; 2) categorising distinct types of ambiguous words, and 3) selecting a matched set of unambiguous words for comparison purposes. In order to aid researchers to address these issues, the aim of the present study is to provide a database of ambiguous and unambiguous Spanish words. In what follows, we will briefly address the

above-mentioned critical issues, focusing on studies that have used homographs (i.e., words with the same spelling but more than one meaning; e.g., *fair*). Then, in the Methods section, we will explain in detail the development of the database.

Selecting an appropriate set of ambiguous words is essential for experimental research on semantic ambiguity. In order to do so, researchers have to determine whether a word is ambiguous or not. This task can be achieved by employing different approaches, either objective or subjective. For example, an objective method to categorize words as ambiguous or unambiguous consists in counting the number of entries of the words in the dictionary (e.g., Jastrzembski, 1981; Rodd et al., 2002). Using this approach (i.e., *the dictionary approach*), words with more than one entry are usually classified as ambiguous, whereas words with only one entry are classified as unambiguous. On the other hand, regarding subjective approaches, a widely used method is to ask participants to generate definitions or lexical associates for a list of words. Then, the definitions or associates are grouped together according to the meaning to which they refer (Nelson, McEvoy, Walling, & Wheeler, 1980; Twilley, Dixon, Taylor, & Clark, 1994). The expected outcome is that definitions or associates for ambiguous words, but not for unambiguous words, will refer to more than one meaning. For instance, in Nelson et al. (1980) some of the associates generated for the ambiguous word *bass* were *fish*, *trout*, *drum, fiddle*, or *guitar*, which are related to their two distinct meanings: *sea perch fish* and *bass guitar*. Finally, another subjective approach is to ask participants to think about all the meanings of a string of letters and then choose a value in a three-point scale: (0) *the word has no meaning*, (1) *the word has one meaning*, or (2) *the word has more than one meaning* (e.g., Kellas, et al., 1988; Pexman, Hino, & Lupker, 2004). With this approach, a quantification of number of meanings (thereafter NOM) is obtained: Words with values close to 2 are classified as ambiguous, and words with values close to 1 are classified as unambiguous.

Each of the above-described methods has some strengths and weaknesses. Concerning objective measures, dictionaries are an exhaustive and standardized resource of words' meanings. Furthermore, most of them list unrelated meanings in different entries, and related meanings under the same entry, facilitating not only the selection of ambiguous and unambiguous words, but also the selection of homonyms (i.e., multiple entries) and polysemes (i.e., only one entry). Given that, dictionaries seem to be a very useful tool to easily select a set of experimental stimuli. However, a strong argument against employing dictionaries in semantic ambiguity research was pointed out by Gernsbacher (1984). She claimed that dictionary definitions should not be taken as a psychologically valid measure of meaning representation. She argued that speakers do not store in their memory all the meanings listed in a dictionary, but only a small sample of them. As an example, she showed that college professors could provide only 2 of the 30 dictionary definitions of the word *gauge*, 3 of the 15 dictionary definitions of the word *fudge*, and 1 of the 15 dictionary definitions of the word *cadet*. Apart from this criticism, another reason against using the dictionary approach is the fact that dictionaries evolve quite slowly (Lin & Ahrens, 2005). Thus, they are plenty of out-dated definitions and do not capture the emerging and gradual changes on the meaning of words.

With respect to subjective measures, asking participants to provide associates or definitions for a set of words would provide a better picture of which meanings are actually represented in the speakers' minds. The reliability of these subjective measures has been evidenced in the studies that have relied on them to select their materials. For instance, it has been demonstrated that associates of ambiguous words produce significant priming effects in LDT (i.e., a facilitation in response time to target words preceded by related [e.g., hit-ball] vs. unrelated [e.g., doctor-ball] prime words; e.g., Klepousniotou, Pike, Steinhauer, & Gracco, 2012). Nevertheless, a significant limitation of using associates is that to determine if a word

is ambiguous or not as well as to count how many different meanings a particular word has, the responses of the participants have to be later examined and categorized by different judges. This task is time consuming, apart from being prone to errors and disagreements, since sometimes it is difficult to determine the meaning to which a particular response refers. For example, it is not clear if the word *land*, produced as an associate of the ambiguous word *plane*, is related to the *airplane* meaning or to the *flat surface* meaning. A similar case occurs when it has to be determined if two definitions provided for a word refer to exactly the same meaning or, instead, if they refer to distinct meanings. Consequently, this limitation may affect the ambiguous/unambiguous categorization of a word and the process of counting the number of different meanings. Finally, asking participants to assess the number of meanings a word has is a straightforward way to known if a word is ambiguous or not for speakers. In addition, it provides a numerical score of the word's ambiguity (i.e., a NOM rating), which facilitates the process of selecting stimuli for an experiment. However, there are two significant weaknesses of this approach. On the one hand, the criterion used by respondents to make their ratings is unknown. On the other, as ambiguous words can widely vary in the number of meanings they have, this approach overlooks the differences between words with few meanings and words with many different meanings (Lin & Ahrens, 2005). Given the above-mentioned advantages and limitations of each approach, we considered convenient to include several different ambiguity measures in the present database, both objective and subjective. Therefore, the database provides the number of dictionary definitions, the number of dictionary senses, NOM ratings, and an ambiguous/unambiguous categorization made by judges on the basis of word associates.

The second critical issue to face when conducting research on semantic ambiguity is how to categorize different types of ambiguous words. These words vary along several dimensions. For example, they vary in the number of meanings (few vs. many meanings), in

meanings' frequency/dominance (balanced vs. unbalanced meanings), and in the degree of relatedness between their meanings (related vs unrelated meanings). Dimensions such as number of meanings or meanings' frequency/dominance have been extensively studied in the past (see Simpson, 1994, for a review); in contrast, relatedness of meanings (hereafter, ROM) has only recently begun to gain experimental interest (see Eddington & Tokowicz, 2015, for a review). Noteworthy, the results from this line of research have suggested that ROM is a relevant variable that determines the experimental effects obtained with ambiguous words.

In order to establish how ROM modulates word processing, a psychologically valid measure of this variable is needed. Researchers have employed different approaches to obtain it. For example, some authors ask participants to rate how related two definitions of an ambiguous word are (Azuma, 1996; Durkin & Manning, 1989). Similarly, other scientists ask participants to judge how related two sentences are, each sentence containing the same ambiguous word but varying the context in which it appears, one context being related to one meaning, and the other related to another meaning (e.g., *Do you have a goal in life? – Liverpool won by goal to nil*; Panman, 1982). These approaches provide a clear context for each word meaning, which forces participants to focus exclusively on a particular meaning when making their decisions. This facilitates the task for the participant and, at the same time, allows researchers to obtain a clean score of the relatedness between the meanings of an ambiguous word. However, the main limitation of these approaches concerns the selection of the meanings to be presented to the participant. If only a subset of all the meanings of an ambiguous word is presented, the ROM rating may be biased (e.g., in case of selecting only two meanings of a word with three meanings, happening that two of them are strongly related but that the third one is unrelated).

An alternative approach to obtain ROM ratings is by asking participants to choose the appropriate ROM value for an ambiguous word using a numerical scale (Hino et al., 2006). In

this procedure, participants are presented with a list of ambiguous words, each one followed by a numerical scale comprising values from *unrelated meanings* to *related meanings*. Then, participants are required to think of all the meanings of each ambiguous word and to judge the relatedness of these meanings by selecting one of the values of the scale. The advantages and limitations of this approach are quite similar to those of NOM ratings. Its main advantages are that it is a straightforward approach to know if a word is polyseme or homonym, and that it provides a numerical value, which could be useful for selecting stimuli for psychological experiments. Regarding its limitations, two should be highlighted: the criterion used by participants to make a ROM decision is unknown, and it is questionable that participants consider all the meanings they know for a word before making the decision. Taking into account the strengths and weaknesses of the different ROM approaches, we decided to include two different ROM measures in the database, each one obtained through a distinct method.

Finally, as previously stated, the third critical issue when conducting research on semantic ambiguity is to choose a proper set of unambiguous words for comparison purposes. This is especially crucial for word recognition experiments, in which researchers typically compare ambiguous and unambiguous words with respect to the time needed by participants to recognize them. As pointed out above, a common finding in LDT is that ambiguous words are recognized faster than unambiguous words (e.g., Hino & Lupker, 1996; Jastrzembski, 1981; Lin & Ahrens, 2010; Millis & Button, 1989). Thus, in order to ensure that this advantage is produced by words' ambiguity, ambiguous and unambiguous words have to be matched on all the relevant variables that are known to affect word processing (e.g., word frequency, word length, familiarity, concreteness, etc.). This experimental control becomes even more important taking into account that several interactions between ambiguity and other variables have been reported. Namely, the ambiguity advantage has been observed for

low frequency ambiguous words, but not for high frequency ambiguous words (e.g., Pexman et al., 2004); for abstract, but not for concrete words (e.g., Tokowicz & Kroll, 2007), and for neutral, but not for emotional words (Syssau & Laxén, 2012). For that reason, in the present database we included ratings for several variables, such as emotional valence, emotional arousal, concreteness, familiarity and age-of-acquisiton.

To sum up, researchers should consider several issues before conducting semantic ambiguity experiments. To address these issues, it is very relevant to have large sets of normed stimuli. There are many databases for English ambiguous words (Azuma, 1996; Durkin & Manning, 1989; Ferraro & Kellas, 1990; Gawlick-Grendell & Woltz, 1994; Gee & Harris, 2010; Gorfein, Viviani, & Leddo, 1982; Griffin, 1999; Nelson et al., 1980; Nickerson & Cartwright, 1984; Panman, 1982; Twilley et al., 1994; Wollen, Cox, Coahran, Shea, & Kirby, 1980). However, the number of available norms in other languages is much lower. Focusing on Spanish, only three normative studies have been published to date (Domínguez, Cuetos, & de Vega, 2001; Estevez, 1991; Gómez-Veiga, Carriedo López, Rucián Gallego, & Vila Cháves, 2010). These databases are described in detail in what follows.

The norms of Dominguez et al. (2001) consist of 100 polysemous words. In this study, a sample of undergraduate students were asked to retrieve the different meanings of each word and to try to write three sentences with them. After that, the authors categorized the sentences according to the meaning to which the ambiguous word referred, and then they counted the number of distinct meanings. The database includes, for each word, the number of meanings, the number of responses assigned to each meaning, and the number of dictionary definitions. On the other hand, the database of Estevez (1991) consists of 152 homonymous words and 61 polysemous words. A sample of 104 participants was asked to write definitions for the homonyms, and 96 participants were asked to write sentences referring to the different meanings of the polysemes. All of them were undergraduate students

living in the Canary Islands. The responses were categorized according to the most common meanings listed for each word in the Spanish Language Dictionary published by the Real Academia Española (RAE, Edition of 1984). The dataset provides, for each word, the number of meanings, the percentage of responses for each meaning, and the number of dictionary definitions. Finally, the norms of Gomez-Veiga et al. (2010) contain 113 ambiguous words, which were evaluated by adults (ranging from 19 to 52 years old) and children (10 and 11 years old). Participants were asked to write down all the meanings they knew for each word in the same order that they retrieved them. The database includes the number of meanings of each word, the percentage of responses assigned to each meaning, and the familiarity of the word.

In light of the above, our opinion is that the available Spanish databases of ambiguous words lack some relevant information. Namely: 1) they only include one subjective measure of ambiguity; 2) they do not provide subjective ROM measures, and 3) none of them has a set of unambiguous words. Thus, the aim of the present study was to construct a normative database of ambiguous and unambiguous Spanish words that overcomes some of the limitations of the available Spanish databases. The present database is made up of 530 words. Four measures of ambiguity were obtained: ambiguity categorization based on lexical associates, subjective ratings on NOM, number of dictionary entries and number of dictionary senses. In addition, we collected two subjective ROM measures that can aid researchers in selecting words to study the polysemy/homonymy distinction. Finally, subjective ratings for some relevant variables (e.g., concreteness, familiarity) known to affect word processing were obtained in order to help researchers with the control of potentially confounding variables in ambiguity studies.

Method

Overview of the procedure used to construct the database

The database was developed following a series of steps. First, 641 Spanish words were selected from the Spanish Language Dictionary published by the Real Academia Española (RAE) (2014). Next, these words were classified as ambiguous or unambiguous according to the associates generated by a group of participants. This classification was first made by four judges (the authors) and then validated by NOM ratings obtained from a different set of participants. After that, two distinct ROM measures for the ambiguous words were obtained. Finally, ratings of concreteness, familiarity, emotional valence, arousal and subjective age-of-acquisition were collected. In what follows, we will explain in detail each step.

Participants

A total of 1213 undergraduate students from the Universitat Rovira i Virgili (mean age = 20.67 years, *SD* = 4.79; 933 females and 280 males) participated in the study in exchange for academic credits. All were highly fluent native speakers of Spanish. Each participant filled in a set of questionnaires.

Item selection

We used an automated script for randomly selecting approximately 1,000 words from an electronic version of the RAE dictionary. This set of words was selected according to some criteria; namely, they should be "potentially unambiguous" or "potentially ambiguous". Words were considered to be potentially unambiguous if they had only one dictionary entry and five or less dictionary senses (e.g., *aeropuerto* 'airport': 1 entry and 1 sense), and (2) words were considered to be potentially ambiguous if they had either more than one dictionary entry or more than five dictionary senses (e.g., *verso* 'verse': 2 entries and 4 senses; *perfil* 'profile': 1 entry and 10 senses). The number of senses for each word was computed as the total number of definitions listed in all the dictionary entries of the word. After removing words with very low frequency, the initial stimuli set consisted of 641 words: 392 potentially ambiguous words and 249 potentially unambiguous words. Words in the set had a lexical

frequency ranging from 0.02 to 2125 occurrences per million ($M = 59.59$, $SD = 175.79$; Duchon, Perea, Sebastián-Gallés, Martí, & Carreiras, 2013), their length was between 3 and 14 letters ($M = 6.46$, $SD = 2.03$), and the most common part of speech was noun (83.3%), followed by verb (7.6%) and adjective (7%). On the other hand, the number of dictionary entries of the words was between 1 and 6 ($M = 1.23$, $SD = 0.60$), and their number of senses ranged from 1 to 49 ($M = 8.41$, $SD = 7.01$).

Procedure

Measures of ambiguity. Following previous studies (e.g., Nelson et al., 1980; Twilley et al., 1994), we opted for collecting participants' responses in a free association task as a subjective measure of ambiguity. All the words from the stimuli set were randomized and listed in 9 questionnaire versions. Then, 236 participants (mean age = 20.4 years, $SD = 4.15$, range = 17 – 47 years; 46 males) were asked to complete a free association task, in which they had to write down the first word that came to mind after reading a cue word. Based on the associates generated by the participants, four judges (the authors) categorized individually each word of the stimuli set as ambiguous or unambiguous. Words were classified as ambiguous if their associates were related to distinct meanings (e.g., for the word *mouse,* participants produced the following associates: *computer*, *cheese*, *mickey*, *cat*, *hamster*, *and keyboard*). Instead, words were classified as unambiguous if all their associates were related to the same meaning (e.g., for the word *major,* participants produced the following associates: *council*, *town*, *politician*, *and president*). Idiosyncratic responses were removed. It is important to note that only words for which full consensus was reached between the four judges were selected; consequently, 111 words that did not fulfil this criterion were removed from the stimuli set. Altogether, of the remaining 530 words, 386 were categorized as ambiguous and 144 as unambiguous.

In order to test the validity of the judges' categorization, and to provide another measure of ambiguity, we collected NOM ratings following the approach used in previous studies (e.g., Ferraro & Kellas, 1990; Pexman et al., 2004). The 530 words were randomized and listed in 10 questionnaire versions. Additionally, 27 nonwords (e.g., *bresio*) were included as fillers in each version. Then, a group of 235 participants (mean age = 21.2 years, SD = 5.53, range = 18 – 57 years; 44 males) was asked to decide if a given character string had *no meaning* (coded as 0), *one meaning* (coded as 1), or *more than one meaning* (coded as 2) using a 3-point scale.

ROM measures. After classifying the words as ambiguous or unambiguous, the next step was to obtain an index of relatedness between the different meanings of each ambiguous word. To this end, we collected two different ROM measures (hereafter, ROM_1 and ROM_2). A novel approach was used to obtain the ROM_1 measure. First, two of the associates generated by the participants were selected for each ambiguous word of the stimuli set. One of the associates was the one with the highest response frequency (hereafter, *modifier1*), and the other was the associate related to a different meaning than that for modifier1 with the highest response frequency (hereafter, *modifier2*). For example, the two associates selected for the ambiguous word *siren* were *sea* (the associate of *siren* with the highest number of responses) and *ambulance* (the associate of *siren* related to a meaning different from that of *sea* with the highest number of responses). With these materials, a questionnaire was constructed. In each page of the questionnaire, the ambiguous word was paired in one line with *modifier1,* and in another line with *modifier2* (e.g., *siren-sea* and *siren-ambulance*) followed by a 9-point scale, ranging from 1 (*unrelated meanings*) to 9 (*same meaning*). Participants were asked to rate to which degree the meaning of the ambiguous word paired with *modifier1* (e.g., in *SIREN-sea*, the meaning *sea nymph*) and the meaning of the ambiguous word paired with *modifier2* (e.g., in *SIREN-ambulance*, the meaning *warning*

alarm) were related. It is important to note that we also included unambiguous words (e.g., *umbrella*), paired with the two associates with the highest response frequency (e.g., *rain* and *water*), with the aim of obtaining an additional validity measure of the judges' categorization.

All the words from the stimuli set (i.e., 386 ambiguous and 144 unambiguous words) were randomized and listed in 8 questionnaire versions [1]. Each version consisted of, approximately, two thirds of ambiguous words and one third of unambiguous words. Detailed instructions and examples were included at the beginning of each version (see the Appendix). One hundred and eighty-three participants (mean age = 19.84 years, SD = 5.52, range = 18 – 68 years; 68 males) completed the questionnaires. Finally, the ROM_1 rating for each word was calculated by averaging the scores of the participants who had assessed it.

In addition, we collected a second ROM measure (ROM_2) by using the same approach used in a previous study (Hino et al., 2006). We randomized and listed the 386 ambiguous words in 9 questionnaire versions, and then we asked 215 participants (mean age = 20.65 years, SD = 5.08, range = 18 – 56 years; 31 males) to think of all the meanings of each word and then to rate the relatedness of the meanings by choosing the appropriate ROM value on a 7-point scale, which ranged from *unrelated meanings* (1) to *highly related meanings* (7).

Other variables. We also collected values of concreteness, familiarity, emotional valence, arousal and subjective age-of-acquisition for the 530 words of the database from different

---

[1] Before administering the ROM_1 questionnaire, we assessed whether participants could easily understand the instructions of the task. To accomplish this, a group of 21 students was asked to complete a pilot version of the questionnaire. The questionnaire consisted of 15 ambiguous and 15 unambiguous words selected from the stimuli set. We calculated the reliability of participants' scores by correlating each participant's data with the mean of the whole sample. The results showed a very high correlation, $r(19)$ = .81, p < .001. In light of these results, we used exactly the same instructions in the final rating study.

sources. Familiarity and concreteness ratings for 412 words were obtained from EsPal (Duchon et al., 2013). Given that the values of concreteness and familiarity for the remaining 118 words were not included in EsPal, they were provided by an additional group of participants. Regarding concreteness, the 118 words were randomized and listed in 2 questionnaires. Forty-four participants (mean age = 22.11 years, SD = 6.31, range = 18 – 56 years; 12 males) rated the words using a 7-point scale, which ranged from 1 (minimum level of concreteness) to 7 (maximum level of concreteness). With respect to familiarity, the 118 words were randomized and listed in 3 questionnaires. A group of 73 participants (mean age = 20.44 years, SD = 2.54 range = 17 – 30 years; 19 males) was asked to rate the familiarity of the words using a 7-point scale, which ranged from 1 (not familiar at all) to 7 (very familiar). To ensure comparability between our ratings and those provided by EsPal, we used the same instructions and rating scales.

In a similar way, emotional valence and arousal ratings were taken from different sources. The values for 191 words were obtained from Guasch, Ferré, and Fraga (2015). The values for the remaining 339 words were obtained from an additional sample of respondents, who were presented with the same instructions and rating scales as those used by Guasch et al. (2015). Concerning emotional valence, the 339 words were randomized and listed in 4 questionnaires. Then, 103 participants (mean age = 20.76 years, SD = 2.46, range = 18 – 29 years; 25 males) were asked to rate emotional valence using the self-evaluation valence scale of the Self-Assessment Manikin (SAM; Bradley & Lang, 1994). The SAM provides a sequence of images to help participants to rate words in an affective scale. The scale for the emotional valence dimension ranges from 1 (strongly negative) to 9 (strongly positive). Regarding arousal, the 339 words were randomized and listed in 4 questionnaires. Participants were provided with the arousal scale of the SAM, which ranges from 1 (not

arousing at all) to 9 (strongly arousing). Ninety-four participants (mean age = 21.72 years, *SD* = 4.87, range = 18 – 61 years; 24 males) answered the arousal questionnaires.

Finally, we collected subjective age-of-acquisition (*AoA*) ratings. As for the above-mentioned variables, different sources were used. AoA ratings for 418 words were obtained from Alonso, Fernández, and Díez (2015). The remaining 112 words were listed in a single questionnaire and presented to a group of 30 participants (mean age = 18.8 years, *SD* = 1.06, range = 18 – 22 years; 11 males). We used the same instructions and scale as Alonso et al. (2015). Respondents were asked to estimate the age at which they learned each word by using an 11-point lineal scale. In this scale, a value of 1 indicates an age *lower than 2 years old*, numbers ranging from 2 to 10 indicate learning ages from *2 to 10 years*, and a value of 11 indicates that the learning age for the word was *11 years or older*.

Results

The complete database is available for download as a supplementary material. A summary of the stimulus characteristics of the database is shown in Table 1. In what follows, we will provide some reliability and validity data for the measures included in the database, as well as the main results of the analyses conducted to fully characterize the normed words.

<Insert Table 1 about here>

Data trimming

All the questionnaire versions for each measured variable were administered to at least 25 participants. Some respondents were rejected after applying a trimming procedure. To do so, we examined the data from the questionnaires in order to identify random or aberrant response patterns. For example, using graphical procedures for assessing person-fit, we discarded those participants with a pattern of responses with almost no variation, as this suggests that they used the same value of the scale for nearly all the words. Furthermore, we also computed a personal correlation coefficient between each participant's data and the

mean, and we excluded those participants with negative values or values near to zero. Values near to zero would suggest that the participant responded randomly to the questionnaire, whereas negative values would suggest that the participant understood the scale in the opposite direction. As a result of applying this trimming procedure, the number of valid responses for each variable was, at least, 21 for NOM, 19 for ROM_1, 19 for ROM_2, 23 for emotional valence, 22 for emotional arousal, 24 for familiarity, and 22 for concreteness.

Reliability

We assessed the reliability of each variable through a split half intergroup procedure. This method consists of dividing the data from each questionnaire into two sets of scores, one set from even items and the other set from odd items. Then, the correlation between the two sets is calculated, giving an index of agreement between participants' scores. Correlations were corrected using the Spearman-Brown formula. Mean intergroup correlation values were $r =$ .85 for NOM ratings, $r = .91$ for ROM_1, $r = .81$ for ROM_2, $r = .84$ for emotional valence, $r =$ .96 for emotional arousal, $r = .98$ for familiarity, $r = .85$ for concreteness, and $r = .88$ for age-of-acquisition. All the correlations were positive and significant (all $p_s < .001$), supporting the reliability of the data.

Relationship between ambiguity measures

We obtained two objective measures of ambiguity (number of dictionary entries and number of dictionary senses) and two subjective measures of ambiguity (NOM ratings and the judges' ambiguous/unambiguous categorization based on word associates). As can be seen in Table 2, objective and subjective measures were positively and significantly correlated. This is of relevance as it supports the validity of the subjective measures that we collected.

<center><Insert Table 2 about here></center>

In addition, it is important to note that ambiguous and unambiguous words categorized by judges differed in dictionary entries, $t(528) = 4.82$, $p < .001$, dictionary senses, $t(528) = 10.76$, $p < .001$, and NOM ratings, $t(528) = 25.79$, $p < .001$.

Relationship between ROM measures

Two ROM measures were collected: ROM_1 and ROM_2. ROM_1 ratings were obtained using a novel approach, whereas to get ROM_2 ratings we followed the procedure used in a previous study (Hino et al., 2006). Importantly, the validity of our novel approach was supported by the high correlation between ROM_1 and ROM_2, $r(528) = .60$, $p < .001$.

Moreover, it should be noted that ROM_1 ratings were not only requested for ambiguous words, but also for unambiguous words. Unambiguous words were included in ROM_1 questionnaires with the aim of obtaining additional support for the judges' ambiguous/unambiguous categorization. Thus, we expected lower ROM_1 values (i.e., closer to the "unrelated meanings" value) for words classified as ambiguous than for words classified as unambiguous. Accordingly, ambiguous words had significantly lower ROM_1 ratings ($M = 3.10$) than unambiguous words ($M = 7.19$), $t(528) = 29.69$, $p < .001$. Likewise, this explains the negative correlation between ROM_1 and the ambiguity measures (see Table 2). In addition, ROM_2 was negatively correlated with both NOM ratings and number of dictionary meanings. As ROM_2 ratings were only obtained for ambiguous words, this negative correlation indicates that ambiguous words with related meanings were considered by participants as having a lower number of meanings than ambiguous words with unrelated meanings.

Comparisons with other Spanish databases of ambiguous words

To assess the validity of the present database, we compared our ratings with those from other Spanish databases of ambiguity (Domínguez et al., 2001; Estevez, 1991; Gómez-Veiga et al., 2010), focusing on overlapping words across databases (ranging from 34 to 78 words). Since

such norms only provide one subjective measure of ambiguity (i.e., number of meanings of the word) and none ROM measure, we could only examine the correlations between our NOM ratings and their measure of number of meanings. Significant correlations were observed between our data and those from Domínguez et al. (2001), $r(32) = .46$, $p = .007$, as well as between our data and those from Gómez-Veiga et al. (2010), $r(57) = .39$, $p = .003$. However, our ratings did not correlate with those of Estevez (1991), $r(76) = .09$, $p = .44$. This may be due to methodological differences between studies. In the norms of Estévez (1991), unlike in the other two studies, the responses of the participants were categorized according to the most common meanings provided by the dictionary. This is of relevance, as it might be possible that some responses related to other meanings were ignored, resulting in a bias in counting the meanings. In addition, Estevez relied on the 1984 edition of the RAE dictionary to categorize participants' responses. There is a difference of 30 years between that version and the current RAE edition. Thus, it is likely that the most common meanings provided by the 1984 edition are slightly different from those included in the current edition. Finally, as the meanings of the words may evolve by the speakers' use of language, it is also likely that the responses of participants in 1991 were somewhat different from those provided by respondents in the present study.

Conclusion

The aim of this study was to provide normative data for a set of ambiguous and unambiguous Spanish words. The resulting database is made up of 530 Spanish words rated on several subjective measures of ambiguity. It also contains dictionary measures of ambiguity (number of dictionary entries and number of dictionary senses), ROM measures and values of some relevant lexico-semantic variables (i.e., concreteness, familiarity, emotional valence, arousal and subjective age-of-acquisition).

The database includes two subjective measures of ambiguity, and two subjective ROM measures. Although the two measures of ambiguity were obtained using distinct approaches (judges' categorization based on word associates and NOM ratings), they were highly intercorrelated. In addition, both measures were correlated with number of senses as well as with number of entries in the dictionary. Furthermore, some significant correlations were observed between our NOM ratings and the ratings from other Spanish databases of ambiguous words. Altogether, these correlations provide support for the validity of the subjective measures of ambiguity. Similarly, ROM measures were obtained using distinct methods. A novel approach was used for collecting ROM_1 ratings, whereas ROM_2 ratings were obtained following the method employed in a previous study (Hino et al., 2006). Despite the differences between both methods, the two ROM measures were highly intercorrelated, supporting the validity of the novel method. In light of this evidence, we consider that both ambiguity and ROM measures could be reliably used in semantic ambiguity research. In particular, they may be of value to assess the influence of the number of meanings and of the relatedness of meanings on word recognition, helping to refine models of word recognition and to test their hypothesis about the processing and representation of ambiguous words.

Furthermore, the database includes some subjective ratings for concreteness, familiarity, emotional valence, arousal and subjective age-of-acquisition that are not provided by any other available Spanish database. Namely, we collected new concreteness and familiarity ratings for 118 words, new emotional valence and arousal ratings for 339 words, and new age-of-acquisition ratings for 112 words. Given that, the present norms will be useful not only for researchers interested in studying semantic ambiguity, but also for researchers looking for controlling such variables in their experiments.

On the other hand, it should be highlighted that the present database might have some applications beyond psycholinguistic research. Indeed, ambiguous words have been used, among other situations, to study clinical populations. For example, by presenting homographs with both personal and impersonal meanings (e.g., *close*, *console*), Hertel & El-Messidi (2006) examined the tendency of people in depressed or dysphoric states to interpret ambiguous events as personally relevant. Similarly, semantic ambiguity has been employed as a tool to investigate the negative interpretation bias related to anxiety disorders. This has been done by examining the response of people with anxiety disorders when they were presented with ambiguous words having both threat and non-threat meanings (e.g., *patient*) (Hayes, Hirsch, Krebs, & Mathews, 2010). A further application can be found in the study of deficits in contextual processing in children with authism. For instance, Hala et al. (2007) investigated whether children with autism are able to make use of the meaning of semantic primes to interpret the meaning of ambiguous words (e.g., pencil-lead). Thus, researchers interested in those or similar applications of semantic ambiguity could be benefited from the present database.

In conclusion, this normative study overcomes some of the limitations of the published databases of Spanish ambiguous words: the scarcity of measures of ambiguity, the lack of subjective ROM measures, and the absence of a set of unambiguous words. As such, the database will be useful for researchers interested in studying semantic ambiguity as well as in their applications in different situations and/or populations. Especially, it will be valuable to assist researchers in categorizing ambiguous and unambiguous words, in categorizing ambiguous words that differ in the relatedness of their meanings, and in preventing any experimental confound due to uncontrolled variables.

References

Alonso, M. A., Fernández, A., & Díez, E. (2015). Subjective age-of-acquisition norms for 7,039 Spanish words. *Behavior Research Methods*, *47*(1), 268–274. doi:10.3758/s13428-014-0454-2

Armstrong, B. C., & Plaut, D. C. (2011). Inducing homonymy effects via stimulus quality and (not) nonword difficulty: Implications for models of semantic ambiguity and word recognition. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2223-2228).

Azuma, T. (1996). Familiarity and relatedness of word meanings: Ratings for 110 homographs. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 109–124. doi:10.3758/BF03203645

Azuma, T., & Van Orden, G. C. (1997). Why SAFE is better than fAST: The relatedness of a word's meanings affects lexical decision times. *Journal of Memory and Language, 36*(4), 484–504. http://doi.org/10.1006/jmla.1997.2502

Balota, D. A., Ferraro, F. R., & Connor, L. T. (1991). On the early influence of meaning in word recognition: A review of the literature. In P. J. Schwanenflugel (Ed.), *The Psychology of Word Meanings* (pp. 187–221). Hillsdale, NJ: Erlbaum.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. doi:10.1016/0005-7916(94)90063-9

Britton, B. K. (1978). Lexical ambiguity of words used in english text. *Behavior Research Methods & Instrumentation*, *10*(1), 1–7. doi:10.3758/BF03205079

Domínguez, A., Cuetos, F., & de Vega, M. (2001). 100 Palabras polisémicas con sus acepciones. *Revista Electrónica de Metodología Aplicada*, *6*(2), 63–84.

Duchon, A., Perea, M., Sebastián, N., Martí, M. A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, *45*(4), 1246–58. doi:10.3758/s13428-013-0326-1

Durkin, K., & Manning, J. (1989). Polysemy and the subjective lexicon: Semantic relatednes and the salience of interword senses. *Journal of Psycholinguistic Research*, *18*(6), 577–612.

Eddington, C., & Tokowicz, N. (2015). How meaning similarity influences ambiguous word processing: the current state of the literature. *Psychonomic Bulletin & Review*, *22*(1), 13–37. doi:10.3758/s13423-014-0665-7

Eich, E. (1984). Memory for unattended events: Remembering with and without awareness. *Memory & Cognition*, *12*(2), 105–111. doi:10.3758/BF03198423

Estevez, A. (1991). Estudio normativo sobre ambigüedad en castellano. *Psicológica*, *3*(2), 237–271.

Ferraro, F., & Kellas, G. (1990). Normative data for number of word meanings. *Behavior Research Methods, Instruments, & Computers*, *22*(6), 491–498. doi:10.3758/BF03204432

Gawlick-Grendell, L. A., & Woltz, D. J. (1994). Meaning dominance norms for 120 homographs. *Behavior Research Methods, Instruments, & Computers*, *26*(1), 5–25. doi:10.3758/BF03204557

Gee, N. R., & Harris, S. L. (2010). Homograph norms: an alternative approach to determining meaning dominance. *Behavior Research Methods*, *42*(4), 976–986.

doi:10.3758/BRM.42.4.976

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical

familiarity and orthography, concreteness, and polysemy. *Journal of Experimental*

Psychology: General, *113*(2), 256–281. doi:10.1037/0096-3445.113.2.256

Gómez Veiga, I., Carriedo López, N., Rucián Gallego, M., & Vila Cháves, J. O. (2010).

Estudio normativo de ambigüedad léxica en castellano, en niños y en adultos.

Psicológica, *31*(1), 25–47.

Gorfein, D. S., Viviani, J. M., & Leddo, J. (1982). Norms as a tool for the study of

homography. *Memory & Cognition*, *10*(5), 503–509. doi:10.3758/BF03197654

Griffin, Z. M. (1999). Frequency of meaning use for ambiguous and unambiguous words.

Behavior Research Methods, Instruments, & Computers, *31*(3), 520–530.

Guasch, M., Ferré, P., & Fraga, I. (2015). Spanish norms for affective and lexico-semantic

variables for 1,400 words. *Behavior Research Methods*. *Advance online publication.*

doi:10.3758/s13428-015-0684-y

Hala, S., Pexman, P. M., & Glenwright, M. (2007). Priming the meaning of homographs in

typically developing children and children with autism. *Journal of Autism and*

Developmental Disorders, *37*(2), 329–40. doi:10.1007/s10803-006-0162-6

Hayes, S., Hirsch, C. R., Krebs, G., & Mathews, A. (2010). The effects of modifying

interpretation bias on worry in generalized anxiety disorder. *Behaviour Research and*

Therapy, 48(3), 171-178.

Hertel, P. T., & El-Messidi, L. (2006). Am I Blue? Depressed mood and the consequences of

self-focus for the interpretation and recall of ambiguous words. *Behavior Therapy*,

37(3), 259–268. doi:10.1016/j.beth.2006.01.003

Hino, Y., Kusunose, Y., & Lupker, S. J. (2010). The relatedness-of-meaning effect for

    ambiguous words in lexical-decision tasks: when does relatedness matter? *Canadian*

    Journal of Experimental Psychology, *64*(3), 180–196. doi:10.1037/a0020475

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An

    alternative to lexical access accounts. *Journal of Experimental Psychology: Human*

    Perception and Performance, *22*(6), 1331–1356. doi:10.1037/0096-1523.22.6.1331

Hino, Y., Lupker, S. J., & Pexman, P. M. (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. doi:10.1037/0278-7393.28.4.686

Hino, Y., Pexman, P. M., & Lupker, S. J. (2006). Ambiguity and relatedness effects in

    semantic tasks: Are they due to semantic coding? *Journal of Memory and Language*,

    55(2), 247–273. doi:10.1016/j.jml.2006.04.001

Hutchison, K. A., & Balota, D. A. (2005). Decoupling semantic and associative information

    in false memories: Explorations with semantically ambiguous and unambiguous critical

    lures. *Journal of Memory and Language*, *52*(1), 1–28. doi:10.1016/j.jml.2004.08.003

Jastrzembski, J. E. (1981). Multiple meanings, number of related meanings, frequency of

    occurrence, and the lexicon. *Cognitive Psychology*, *13*(2), 278–305. doi:10.1016/0010-

    0285(81)90011-6

Kellas, G., Ferraro, F. R., & Simpson, G. B. (1988). Lexical ambiguity and the timecourse of

    attentional allocation in word recognition. *Journal of Experimental Psychology. Human*

    Perception and Performance, *14*(4), 601–609. doi:10.1037/0096-1523.14.4.601

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in

    word recognition: An advantage for polysemous but not homonymous words. *Journal of*

Neurolinguistics, *20*(1), 1–24. doi:10.1016/j.jneuroling.2006.02.001

Klepousniotou, E., Pike, G. B., Steinhauer, K., & Gracco, V. (2012). Not all ambiguous

words are created equal: An EEG investigation of homonymy and polysemy. *Brain and

Language, 123*(1), 11–21. http://doi.org/10.1016/j.bandl.2012.06.007

Lin, C-J. C., & Ahrens, K. (2005). How many meanings does a word have? Meaning

estimation in Chinese and English. In J. W. Minett & W. S.-Y. Wang (Eds.), *Language

acquisition, change and emergence: Essays in evolutionary linguistics* (pp. 437–464).

Hong Kong: City University of Hong Kong Press.

Lin, C-J. C., & Ahrens, K. (2010). Ambiguity advantage revisited: Two meanings are better

than one when accessing Chinese nouns. *Journal of Psycholinguistic Research*, *39*(1),

1–19. doi:10.1007/s10936-009-9120-8

Mathews, A., Richards, A., & Eysenck, M. (1989). Interpretation of homophones related to

threat in anxiety states. *Journal of Abnormal Psychology*, *98*(1), 31–34.

doi:10.1037/0021-843X.98.1.31

Millis, M. L., & Button, S. B. (1989). The effect of polysemy on lexical decision time: now

you see it, now you don't. *Memory & Cognition*, *17*(2), 141–147.

doi:10.3758/BF03197064

Nelson, D. L., McEvoy, C. L., Walling, J. R., & Wheeler, J. W. (1980). The University of

South Florida homograph norms. *Behavior Research Methods & Instrumentation*, *12*(1),

16–37. doi:10.3758/BF03208320

Nickerson, C. A., & Cartwright, D. S. (1984). The University of Colorado meaning norms.

Behavior Research Methods, Instruments, & Computers, *16*(4), 355–382.

doi:10.3758/BF03202462

Panman, O. (1982). Homonymy and polysemy. *Lingua*, *58*(1-2), 105–136. doi:10.1016/0024-3841(82)90059-6

Pexman, P. M., Hino, Y., & Lupker, S. J. (2004). Semantic ambiguity and the process of generating meaning from print. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1252–1270. doi:10.1037/0278-7393.30.6.1252

Real Academia Española. (2014). Diccionario de la lengua española (23rd ed.). Madrid, Spain.

Rodd, J. M., 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

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28*(1), 89–104. http://doi.org/10.1016/j.cogsci.2003.08.002

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9*(5), 487–494. http://doi.org/10.1016/S0022-5371(70)80091-3

Simpson, G. B. (1994). Context and the processing of ambiguous words. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 359–374). San Diego, CA: Academic Press.

Swinney, D. A. (1991). The resolution of interdeterminacy during language comprehension: perspectives on modularity in lexical, structural and pragmatic process. In G. B. Simpson (Ed.), *Understanding Word and Sentence* (pp. 367–385). North Holland: Elsevier Science Publishers B.V.

Syssau, A., & Laxén, J. (2012). The influence of semantic richness on the visual recognition of emotional words. *Canadian Journal of Experimental Psychology*, *66*(1), 70–8. doi:10.1037/a0027083

Tokowicz, N., & Kroll, J. F. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes*, *22*(5), 727–779. doi:10.1080/01690960601057068

Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory & Cognition*, *22*(1), 111–126. doi:10.3758/BF03202766

Van Petten, C., & Kutas, M. (1987). Ambiguous words in context: An event-related potential analysis of the time course of meaning activation. *Journal of Memory and Language*, *26*(2), 188–208. doi:10.1016/0749-596X(87)90123-9

Wollen, K. A., Cox, S. D., Coahran, M. M., Shea, D. S., & Kirby, R. F. (1980). Frequency of occurrence and concreteness ratings of homograph meanings. *Behavior Research Methods & Instrumentation*, *12*(1), 8–15. doi:10.3758/BF03208319

Appendix

Instructions for the ROM₁ questionnaire. Translated from the original instructions in
Spanish.

As you may already know, there are different types of ambiguous words. There are
ambiguous words with unrelated meanings, and ambiguous words with related meanings. For
example, *siren* is an ambiguous word with two unrelated meanings: it may refer to *a warning*
device or to *a sea nymph*. On the contrary, the ambiguous word *balloon* has several related
meanings: it may refer to *a rubber bag that can be inflated with gas* or to *a bag of strong,*
light material filled with a gas lighter than air so as to rise through the air.

Below, we present you with a word assessment questionnaire. In this questionnaire, you will
find both ambiguous and unambiguous words. Each word to be assessed will appear together
with other two words, which we will call *modifiers*. Modifiers are used to direct your
attention toward a specific meaning of the first word. Your task consists of comparing the
meaning of the word to be evaluated when it is accompanied by the first modifier, to the
meaning of that word when it is accompanied by the second modifier. In order to make your
ratings, you have to use a scale from 1 to 9, where 1 indicates that the two meanings are
completely different, and 9 that the two meanings are exactly the same. We encourage you to
use all the values on the scale.

Original instructions in Spanish for the ROM₁ questionnaire.

Como ya sabrás, hay distintos tipos de palabras ambiguas. Existen palabras ambiguas con
significados no relacionados y palabras ambiguas con significados relacionados. Por ejemplo,

sirena es una palabra ambigua con significados no relacionados: puede referirse a 1) *sonido de alerta* o a 2) *ninfa del mar*; en cambio, la palabra ambigua *globo* posee significados relacionados: puede referirse a 1) *bolsa de goma o de otro material flexible que se llena de gas o de aire* o a 2) *nave aerostática formada por una bolsa que, llena de un gas de menor densidad que el aire atmosférico, eleva una barquilla sujeta a su parte inferior*.

A continuación te presentamos un cuestionario de evaluación de palabras. En el cuestionario encontrarás tanto palabras ambiguas como palabras no ambiguas. Cada palabra a evaluar aparecerá acompañada por otras dos, a las cuales denominamos *modificadores*. Los modificadores sirven para dirigir tu atención hacia un significado concreto de la primera palabra. Tu tarea consiste en comparar el significado de la palabra a evaluar acompañada por el primer modificador con el significado de la misma acompañada por el segundo modificador. Expresarás tu respuesta en una escala de 1 a 9, donde 1 indica que los significados son completamente distintos, y 9 que se trata de exactamente el mismo significado. Intenta utilizar todos los valores de la escala.

Instructions for the ROM$_2$ questionnaire. Translated from the original instructions in Spanish.

Below you will find a list of ambiguous words, that is, words that have more than one meaning. We ask you to:

1) Read each word (e.g., "siren")

2) Think of all the meanings of that word (e.g., "warning device" and "sea nymph")

3) Rate the relatedness of these meanings

We provide you a scale ranging from 1 ("unrelated meanings") to 7 ("highly related meanings") to make your ratings. You should use low relatedness values if the meanings are not related. For example, you can use a low relatedness value to rate the ambiguous word "siren", since it has two unrelated meanings (i.e., "warning device" and "sea nymph"). Instead, you should use high relatedness values if the meanings are related. For example, you can use a high relatedness value to rate the ambiguous word "newspaper", as it refers to multiple related meanings: 1) a publication, usually issued daily or weekly; 2) a business organization that prints and distributes such a publication; 3) a single issue of such a publication, and 4) the paper on which a newspaper is printed.

Important: Although all the words of the questionnaire are ambiguous, it is possible that initially you only remember one meaning of a given word. If you find yourself in that situation, take as much time as you need to remember all its meanings. If you are still unable to remember any other meaning, you can leave that item unanswered.

Original instructions in Spanish for the ROM$_2$ questionnaire.

A continuación verás una lista formada por palabras ambiguas, es decir, palabras que poseen más de un significado. La tarea que debes realizar consiste en:

1) Leer la palabra (p.ej., "sirena")

2) Pensar en TODOS los significados de esa palabra (p.ej., "alarma" y "ninfa del mar")

3) Indicar si los significados de esa palabra están relacionados entre ellos o no

Para evaluar el grado de relación de los significados dispones de una escala con puntuaciones de 1 a 7. Utiliza puntuaciones bajas si los significados de la palabra no están relacionados.

Por ejemplo, la palabra ambigua "sirena" deberías puntuarla con un valor bajo, ya que posee dos significados no relacionados ("sirena" y "ninfa del mar"). En cambio, utiliza puntuaciones altas si los significados de la palabra están relacionados. Por ejemplo, la palabra ambigua "diario" deberías puntuarla con un valor alto, puesto que tiene significados relacionados entre sí: 1) que se corresponde a cada día (p.ej., desayuno diario) y 2) periódico que se publica todos los días (p.ej., El País, El Mundo, etc.).

Importante: Aunque todas las palabras del cuestionario son ambiguas, es posible que en un primer instante sólo recuerdes un significado de alguna(s) de ellas. Si te encuentras en esa situación, tómate todo el tiempo que sea necesario para intentar recordar todos sus significados. Si aun así te es completamente imposible recordar ningún otro significado, puedes dejar la respuesta en blanco.

Table 1

Descriptive stimulus characteristics for the 386 ambiguous words and the 144 unambiguous words (classified according to the judges' categorization measure)

	NOM	Entries	Senses	ROM$_1$	ROM$_2$	AoA	Fam	Con	Val	Aro	Freq	Lett
					Ambiguous words							
Mean	1.66	1.30	9.77	3.10	3.66	6.68	5.49	4.39	5.07	4.42	70.5	6.20
SD	0.23	0.66	6.79	1.53	1.18	1.88	0.79	0.94	1.03	0.94	203.8	1.89
					Unambiguous words							
Mean	1.13	1.03	3.38	7.19		6.92	5.31	5.60	4.94	4.37	20.7	7.53
SD	0.15	0.23	3.49	1.01		2.36	1.04	0.88	1.16	0.95	66.2	2.20

Note. Ent = number of dictionary entries; Sen = number of dictionary senses; AoA = Age of acquisition; Fam = familiarity; Con = concreteness; Val = valence; Aro = arousal; Freq = frequency (number of occurrences per million); Lett = word length (in letters).

Table 2

Correlations between the measures of ambiguity and the ROM measures

	Measure	1	2	3	4	5	6
1	Dictionary entries	–					
2	Dictionary senses	.14**	–				
3	Judges' categorization	.21***	.42***	–			
4	NOM	.24***	.50***	.75***	–		
5	ROM$_1$	-.27***	-.39***	-.79***	-.73***	–	
6	ROM$_2$	-.35***	.05	–	-.29***	.60***	–

** *p* < .01, *** *p* < .001