How lexical ambiguity distributes activation to semantic neighbors

Some possible consequences within a computational framework

Guillermo Jorge-Botana and Ricardo Olmos
Universidad Nacional de Educación a Distancia (UNED), Spain /
Universidad Autónoma de Madrid, Spain

The role which the diversity of a word's contexts plays in lexical access is currently the object of research. Vector-space models such as Latent Semantic Analysis (LSA) are useful to examine this role. Having an objective, discrete model of lexical representation allows us to objectify parameters in order to define contextual focalization in a more measurable way. In the first part of our study, we investigate whether certain empirical data on ambiguity can be modeled by means of an exclusively symbolic single representation model such as LSA and an excitatory-inhibitory mechanism such as the Construction-Integration framework. Our observations support the idea that some ambiguity effects could be explained by the contextual distribution using such a model. In the second part, we put abstract and concrete words to the test. Our LSA model (exclusively symbolic) and the excitatory-inhibitory mechanism can also explain the penalty paid by abstract words as they activate other words through semantic similarity and the advantage of concrete words in naming and semantic judgments, though it does not account for the advantage of concrete words in lexical decision tasks. The results of this second part are then discussed within the framework of the embodied/symbolic view of the language process.

Keywords: Representation of Ambiguity, Polysemy, Abstractness, *Contextual Diversity*, Entropy, Vector Space Models, Distributional properties, Latent Semantic Analysis (LSA), Imagery, Second Order Isomorphism

Traditionally, the definition of ambiguity has been formulated in several different ways. For example, lexicographers have attempted to identify the cases in which the usage of a term modulates its meaning or configures a new meaning (see

The Mental Lexicon 9:1 (2014), **67–106.** DOI 10.1075/ml.9.1.04jor
ISSN 1871–1340 / E-ISSN 1871–1375 © John Benjamins Publishing Company

Kilgarriff, 1997). Cognitive linguistics and psycholinguistics have also attempted to show how lexical ambiguity might be represented in the mind. Some models have argued for the existence of different entries for each of the meanings of ambiguous words (Klein & Murphy, 2001), while others hypothesize the existence of a common representation for all meanings and their specific representations (Rodd, Gaskell, & Marslen-Wilson, 2002). But it is the old definition comprising difficulty of association, offered by Brown & Ure (1969), which provides our starting point. Closely related to ambiguity, polysemy seems to be defined by the fact that a representation has several subject area focal points. As a result, the strength of the relationships it might form with other words is reduced. This phenomenon may be related to the experimental data that show disadvantage for ambiguous words in certain tasks. Duffy, Morris, & Rayner (1988) found, for example, that ambiguous words with equally probable meanings that are not biased toward a given context, draw longer fixation times from readers. When the meanings are not equally probable (when there is a dominant meaning) there is no difference in fixation times compared with non-ambiguous words. Moreover, when there is a predominant meaning and the context biases towards the non-predominant meaning, the differences in favor of the non-ambiguous meaning arise again. Thus, polysemous words seem to cause some processing difficulties when retrieving sense in the absence of a predominant context. At the same time, a second effect has been described concerning ambiguous words. Some studies have found that, in the Lexical Decision Task (LDT), ambiguous words require a shorter response time (Hino, Lupker, & Pexman, 2002; Pexman & Lupker, 1999). Many of these studies argue that response times to ambiguous words in LDT are favored by the fact that more semantic entries will be activated from the orthographic representation (a pattern of activation that amalgamates several meanings), and therefore in turn, more orthographic entries (with top-down processing). This non-specific activation is enough to make them reach the decision threshold earlier than in the case of non-ambiguous words (Besner & Joordens, 1995; Joordens & Besner, 1994; Pexman, Hino, & Lupker, 2004).

Given these data, it seems that lexical ambiguity yields contradictory results. On the one hand, it makes reading comprehension slower, but on the other it seems to speed up recognition of a letter string such as a word. There seems to be justification for the term "efficient then inefficient", coined by Piercey & Joordens (2000) in a model that tried to accommodate the dissonant behavior of ambiguous words in the lexical decision task (advantage) and reading comprehension (disadvantage). This model proposes the existence of two differentiated steps when processing ambiguous words: a first step where non-specific activation is generated and a second step where activation moves toward certain meaning patterns. It also assumes that lexical decision is prior to full semantic analysis and does not require a detailed analysis, and also that reading comprehension does require such a semantic analysis. It follows that ambiguous words have an advantage in the LDT by generating more non-specific activation, but that this advantage is lost in reading, as their multiple meanings make the second step take longer, as activation is channeled towards different, separate patterns.

Study I: Distributional Properties and Ambiguity

One of the characteristics that define words is their distributional properties and the type of relationships that they promote with other terms due to these properties. For instance, Adelmam et al. (2006) found that their distributional index, *Contextual Diversity,* neutralized the effects of classical Word Frequency. So distributional properties can have consequences in the two stages of lexical access which were put forward by Piercey & Joordens (in both polysemous and monosemous words): indiscriminate activation of contents in the first stage and meaning hypothesis activation in the second one.

Because of the fact that LSA (Latent Semantic Analysis) is sensitive to the distributional properties of the semantic representation of words and has obtained remarkable results simulating lexical representation in humans (Huetti, Quinlan, McDonald, & Altman, 2005; Kintsch & Mangalath, 2011; Landauer & Dumais, 1997), we argue that an LSA framework might help to throw light on theories of lexical representation. If LSA can proportionally simulate certain empirical phenomena involving lexical ambiguity in a given task, it could be inferred that the functional architecture of the lexical system could act analogously to such a model in some situations, with a single, exclusively symbolic representation. So our question is: how might we account for the "efficiency then inefficiency" scenario displayed in the empirical data using a vector space model as LSA? We might argue that vectorial distribution of terms can explain why ambiguous words are more activated in the first state of lexical access. In turn, taking vector distribution also into account and with the assistance of an activation-inhibition mechanism, it is possible to explain why non-ambiguous words have an advantage with respect to ambiguous words in the second processing phase, when resources focus on finding a meaning hypothesis.

The rows in Figure 1 show seven vectors (indicated by letters) representing seven words. The shaded cells indicate which dimensions (represented by the different columns) are activated for each word. Word A is ambiguous since it scores on four of the dimensions. It therefore promotes similarity with some terms (with vector B, as it shares a strong weight in dimensions 2 and 4), but at the same time this similarity will be penalized by the other shaded cells 6 and 8 (which also

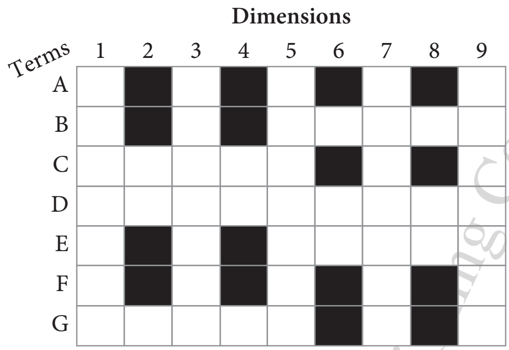

Figure 1. A simplified representation of some kind of words in Vector Space Models, A being an ambiguous word and B, C, D, E, F and G unambiguous words.

relate its meaning to Vector C). The same effect will be found in the dimensions of the other meaning — in other words, it will favor relationships with terms that have this meaning but penalize relationships that have other meanings (A with C). In this way, vectors that represent many meanings will be penalized for taking on synonymous relationships, displaying associative difficulty, which causes semantic neighbors to receive less activation and be more variable as regards content in the first stage, where activation still does not move toward certain meaning patterns by means of inhibition-activation mechanisms (Kintsch, 1998; Piercey & Joordens, 2000). This was reported in an informal LSA study by Wandmacher (2005), who found that the terms with the highest numbers of erroneous relationships seem to occur in many contexts (words such as "think" and "example"). This effect would not be found in unambiguous words, as the dimensions that represent them promote only relationships with a single meaning, without penalizing others (D with E and F with G). So owing to the very distribution of the vectors of ambiguous words, the relationships with their *n* first semantic neighbors (proximal neighbors) are lower in similarity and more spurious than those of non-ambiguous words. In turn, this disadvantage of ambiguous words as regards similarities to their closest neighbors can have damaging consequences when making meaning hypotheses, that is to say, when, in a second stage of the process, after the initial indiscriminate activation (where activation moves toward certain meaning patterns), activation-inhibition mechanisms between the activated contents have been triggered. This becomes particularly relevant when it is assumed that the closest contents are crucial for these mutual constriction mechanisms, as argued by some authors (Kintsch, 1998; Mirman & Magnuson, 2008; Piercey & Joordens, 2000).

At the same time, following the model expressed in Figure 2, ambiguous words have a certain advantage in activating the rest of their neighbors (distal neighbors), as ambiguous vectors represent more content than unambiguous ones, and may therefore fit more easily into the patterns of the other terms. In the first stage of indiscriminate activation, ambiguous words activate the remaining contents in a more intense way, which would be more numerous than the closest content. In short, ambiguous vectors have more difficulty matching the patterns of their proximal neighbors (associative difficulty) but match more closely all others (distal neighbors), so ambiguous vectors would trigger higher and earlier non-specific activation than non-ambiguous vectors. This non-specific activation could be enough to cross the decision threshold to say that it is a word earlier than in a non-ambiguous words in LDT, which does not require detailed analysis, as has been proposed (Piercey & Joordens, 2000). In fact, this distal set of neighbors has been reported in the Mirman & Magnuson (2008) study to have early, facilitative and non-inhibitory effects.

This theoretical framework could be upheld in all linguistic definitions that fall within the definition of ambiguity, something that might include such phenomena as polysemy and homonymy.

Hypotheses

There are three hypotheses in the first study. They all test the behavior of our LSA model to see whether they fit the empirical data. The first hypothesis will establish the difference between distribution of polysemous and monosemous words. The second hypothesis concerns the expected differences between polysemous and monosemous words when it comes to activating their potential neighbors in an indiscriminate, undifferentiated way. These differences might account for the first phase of activation of words between which no inhibitory mechanisms are yet established. The third hypothesis tries to account for the differences between monosemous and polysemous words simulating an activation-inhibition mechanism between the neighbors which have been most activated. This hypothesis would belong to a later processing stage. The hypotheses can be described as follows:

First Hypothesis

The entropy index indirectly measures whether the distribution of a vector is focalized upon a few or many contexts. We hypothesize that, in our model, the vectors of polysemous words have higher entropy than the vectors of monosemous words. It would mean that the vectors of polysemous words are distributed along many contexts. Therefore, the statistical hypothesis is that there is a statistically significant difference between averages in favor of polysemous words in the entropy index. As can be guessed, this may be a minor hypothesis, but it serves to align the model with prior models such as the contextual diversity model and as a starting point for the other arguments in this study.

Second Hypothesis
We offer the hypothesis that this very distribution of the vectors of polysemous words makes them activate their proximal set of neighbors less intensely than monosemous words. However, given the higher non-specificity of polysemous words, they active the remaining semantic neighbors, which are much more numerous, more intensely. Translating this theoretical hypothesis into a statistical hypothesis of our model, we will show that the function that describes the curve of a word's semantic relationship with its semantic neighbors supports such a statement. Thus our hypothesis translates into the curve of polysemous words' similarities with semantic neighbors falling more sharply (significant differences in the function-defining parameters) than that of monosemous words. In this same way, taking the curves in our models, we predicted an advantage of polysemous words when activating the neighbors in the distal zone, which is objectively larger than the proximal zone. Given that all this indiscriminate activation between words and neighbors takes place in a first stage of lexical access, our model might account for the advantage of polysemous words in tasks in which inhibition mechanisms play no role but rather brute activation counts, as we believe is the case of LDT.

Third Hypothesis
We also argue that, given the difference between both words' distributions, polysemous words will have higher costs in our model when making a meaning hypothesis from proximal neighbors. In other words, when these proximal neighbors are subjected to a mutual constriction mechanism, monosemous words will find more solid meaning hypotheses at lower time and costs. In this case, the main statistical hypothesis is that the number of cycles and costs necessary to stabilize monosemous words is lower than the number required to stabilize polysemous words. If the model behaved like this, it might explain the reason for the advantage of monosemous words in tasks which do require inhibition mechanisms triggered, in a second stage of lexical access, as we claim is the case of comprehensive reading and categorization tasks.

Procedure

To verify the hypothesis of distributional properties of vectors as the cause of the effects described in polysemous/monosemous words, we ran the LSA procedure (see Deerwester et al., 1990) on the Spanish corpus Lexesp (Sebastián, Martí, Carreiras, & Cuetos, 2000). Basically, LSA is a vectorial representation of language which is constituted exploiting word occurrences in different contexts. It can be conceived as an automatic sequence of mathematical and computational methods for representing the meaning of words and text passages (Landauer & Dumais, 1997). To obtain this vectorial representation, large linguistic corpora (or text samples) are processed. A key issue is what can be done with LSA. Each term of this corpus is expressed in *k*-dimensional vector. On the one hand it is possible to measure similarities of words. LSA measures the similarity between two words using the cosine between the two vectors that represents the in the *k*-dimensional space. On the other hand, we can analyze the contextual distribution of each word in this mathematical representation.

We generated a sample of polysemous and monosemous words controlled on the frequency, familiarity, concreteness and imageability to be represented in such a space. To objectify the fact that the distributional properties of both types of words are different, we calculated the entropy index. In addition, to emulate the indiscriminate activation in a first stage, their neighbors' lists (by cosines) in the form of ranks were generated. Both curves of ranks, polysemous and monosemous, were adjusted to a theoretical function and were compared. Finally, we applied a mechanism to emulate the excitation-inhibition undergone by each group of words in the final stage. For this purpose, we applied the Construction-Integration algorithm (Jorge-Botana, Olmos, & Barroso, 2012; Kintsch, 1998; Kintsch, 2000; Kintsch, 2001; Kintsch & Bowles, 2002; Millis & Larson, 2008) and measured the costs of reaching a meaning.

Concerning the LexEsp corpus, a "hand-made" lemmatized version was used (where plural forms were transformed into the singular and feminine forms were transformed into masculine, while all verbs were standardized to their infinitive form). We chose sentences as processing units. We deleted words that appeared in less than seven documents. This ensures a minimal representation of the terms analyzed. We then applied entropy pre-processing, finally obtaining a matrix with 18,174 terms in 107,622 documents (sentences), to which we applied the Singular Value Decomposition, reducing the three resulting matrices to 270 dimensions. For this purpose, we used Gallito®, an LSA tool implemented in our research group developed in .Net® (Jorge-Botana, Olmos, & Barroso, 2013).

Method

The procedure for extracting the sample of words to be studied was as follows: we took polysemous words using two sources (Estévez, 1991; Jorge-Botana, León, Olmos, & Hassan-Montero, 2010). Using these two sources, our criteria to select the words for the polysemy group were: (1) that the words should have significantly more entries in the RAE (Spanish Royal Academy) dictionary than monosemous words ($T(46) = 4.85, p < .005$) and (2) that in Lexesp at least two meanings of each polysemous word were represented, checked using the authors' criteria with a visual sample of the first 100 semantic neighbors of each word. In total, 24 polysemous words were chosen. The group of 24 monosemous words was extracted from the Lexesp corpus and matched[1] with the polysemy group on indices of frequency-log10 in the corpus ($T(46) = .24, p = .81$), familiarity ($T(46) = .40, p = .69$), concreteness ($T(46) = .82, p = .42$), and imageability value ($T(46) = 1.71, p = .09$). To extract the values we use BuscaPalabras (Davis & Perea, 2005), a program that automatically assigns the values to each term.

Three analyses were carried out:

First, the entropy index was calculated for each word in the raw occurrence matrix for the Lexesp corpus. This index specifies the amount of information that a word offers in relation to the contexts in which it appears. A higher degree of entropy tends to indicate a greater lack of focal points in terms of subject matter. If a word lacks subject area focalization, its occurrence is not enough to predict the topic of our discourse. We calculate the entropy index with the global weight in Formula 1. A high global weight means low entropy (see Nakov, Popova, & Mateev, 2001). In other words, if a word appears in an extreme number of contexts its global weight is low, we estimate that its entropy is high. We would expect polysemous words, which appear in a greater number of contexts, to have higher entropy than monosemous words. With this prediction in mind, the global weighting of all previous words was calculated.

$$g_i = 1 + \sum_j \left[\frac{p_{ij} \log(p_{ij})}{\log(n)} \right] \qquad \text{(Formula 1) Global weight}$$

Where:

$$p_{ij} = \frac{tf_{ij}}{gf_i}$$

tf_{ij} is the number of occurrences of term i in document j, gf_i is the total number of times term i occurs in all documents, and n is the number of documents.

Secondly, the indiscriminate activation in a first stage of the process was simulated extracting a huge list of neighbors for each word. The first 5000 neighbors of each word in the monosemous and polysemous sets were extracted and an average of each group was taken in a single score per rank (rank 1 = first similar neighbor average cosine, rank 2 = second similar neighbor average cosine, and so on). It is important to note that rank distribution does not denote temporality: that is to say, a neighbor with rank 1 is not activated before rank 556, but simply potentially receives more initial and indiscriminate activation in a first stage of lexical process. We finally obtained two distributions of ranks, one for polysemous words and another for monosemous word. To find out whether the curves for monosemous words are statistically different from those of the polysemous words, the following procedure was followed: each of the curves for every word was adjusted to the Napierian logarithm Function, which had the best fit (Formula 2) and the two parameters of the function (*a* and *b*) were extracted. Using these two parameters as dependent variables, we carried out a t test for independent samples in order to compare the means for the two parameters.

$$f(x) = -a \cdot \log(x) + b \qquad \text{(Formula 2) Logarithm function}$$

In addition, taking the average function that represents monosemous words and a different one for polysemous words and adjusting both to the logarithmic function, analyses were carried out to calculate activation areas and whether a global advantage for polysemous words as described in the hypotheses exists was verified.

Third, to answer the question of what would happen in the proximal area of the distribution if we applied a mechanism to simulate the activation-inhibition when building meaning in a second stage of the lexical access, we applied one of the most widely used algorithms in combination with LSA. This is the C-I (Construction-Integration) algorithm (Jorge-Botana et al., 2012; Kintsch, 1998; Kintsch, 2000; Kintsch, 2001; Kintsch & Bowles, 2002; Millis & Larson, 2008). The C-I mechanism is simple. In a first phase, known as the Construction stage, the first neighbors are extracted from all the terms that constitute a given text. In a second phase, a connectionist network is built using all of them, and is run until a stable point is achieved. In the same way as a text, a term — particularly a polysemous term — also has neighbors which are part of the various meanings represented in it. Therefore, by extracting the neighbors of that term and building the same network with it, we will be acting in an analogous way to a text (both are represented semantically by components in only one vector). The n first neighbors of the term will enter the mutual constriction dynamics until a stable meaning hypothesis is achieved. Technically, the C-I mechanism for a term

T was implemented as follows: in a first phase, the Construction phase, the n first neighbors of the term T were extracted (including itself), constituting a matrix M with all of them (which is n x n), in which every cell represents the cosine between each neighbor (all with all). It is clear that the diagonal is taken up by 1. Moreover, a vector A is instantiated which is n x 1, where the activations of each of the terms in matrix M will be. These activations in A will change in each of the process cycles until a final meaning is defined. In our case, the initial activation of this vector A is not 1 or random as in previous studies, but rather the cosine of term T with each of the terms whose activation is represented in that vector. This accounts for the activation which each neighbor has previously received from each reference term T.

Thus we have an M matrix which will modulate coordinates in the form of neighbors that are activated when the meaning of a term T is accessed. The activations of those neighbors are represented in vector A. From here on, vector A is multiplied by matrix M in each of the cycles, obtaining again a vector (n x 1). This vector is the new activation vector for that cycle. This vector is normalized and multiplied again by M, obtaining another activation vector for that cycle, which is then normalized, and this procedure is repeated until a stable position is reached, which means that a meaning hypothesis has been obtained. This stable position is marked by a stop criterion, which will be met when the differences between the dimensions of activation vector in cycle i – 1 and cycle i are minuscule: in our case, less than 0.000001. This is the final cycle, which designates the number of necessary cycles.

How is the cost of reaching a meaning measured? Some measurements have been proposed. Millis & Larson (2008) employ the number of cycles to try and reach the stable point, as well as the average activation in the final vector. Millis & Larson suggested that they are indicators of the Integration phase of the C-I algorithm. The number of cycles might be an indicator of the time for construct a meaning hypothesis and the final activation values might be an indicator of its coherence with a meaning. In addition to these, we also use the state correlation: The correlation between the initial activation vector (which represents the initial activation for every neighbor) with the final activation vector. The lower this correlation is, the larger the differences between the initial and the final vector will be. Therefore, the vectors are different and this entails a larger effort to reach a final meaning hypothesis. Of these three indicators, the number of cycles gives the most reliable information because it more directly maps onto Reaction Times, but we preferred to put all of them in to the test. Following this method, we analyzed the critical rank proximal area of the distribution, where we claim differences are found between polysemous and monosemous words.

An important question is what the *n* of this critical rank with which matrices M and A of term T should be. Two criteria are envisaged to select the number of neighbors that will be part of the network and thus part of this analysis. The first criterion is assigned by rank, ranks being 100, 200, 300 and 400; that is to say, neighbors are extracted up to rank 100, 200, and so on. This first criterion is the one used in the original conception of the algorithm (Kintsch, 1998; Kintsch & Patel, 1999; Kintsch, 2000; Kintsch, 2001; Kintsch & Bowles, 2002). In these simulations, n is established a priori. The other criterion is an activation criterion, and neighbors are extracted until a certain activation level is reached. This criterion is the one used by Mirman & Magnuson (2008) in their "attractor model", which sets 0.25, 0.5 and 0.75 as thresholds. In our case, the critical area in rank distribution recommends using 0.2 and 0.3 cosines as thresholds.

Results

First Hypothesis: Global Weight

The results (Figure 2) show that entropy is higher for polysemous words (they have a lower global weight) than in monosemous words ($T(46) = 2.14$, $p = .038$) a finding that fits earlier results. According to this, polysemous words in our model

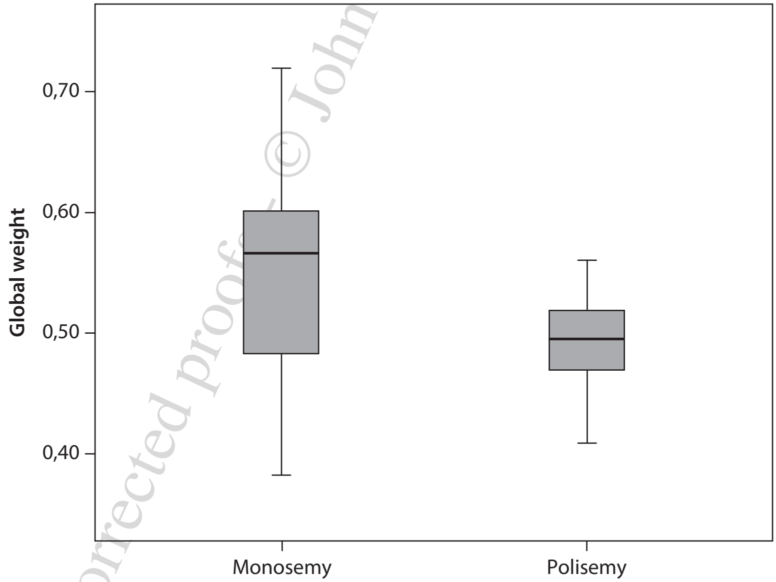

Figure 2. Means global weights for polysemous and monosemous words. Note that the more global weight, the less entropy.

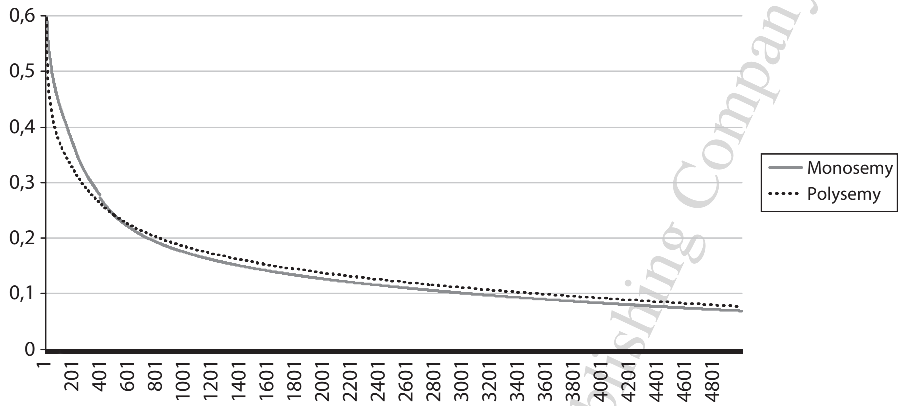

Figure 3. Distributions of ranks and the average cosine, one for polysemous words and another for monosemous words. The horizontal axis represents the first 5000 neighbors sorted by the average cosine of each condition. The vertical axis represents the cosines. For instance, the 1st rank is the closest term and corresponds to an average cosine above 0.6. The 1001st rank corresponds to a cosine lower than 0.2.

have a greater tendency towards dispersal of contextual topics. Figure 3 shows global weight distributions for both word groups.

Second Hypothesis: Rank Distribution for the First 5000 Neighbors
To represent rank distribution, we calculated the cosine of every term T with respect to all the other corpus terms, and chose the 5,000 terms with the highest cosines. These are the 5,000 semantic neighbors closest to each term. The graph in Figure 4 shows, in descending order, the average cosines of the 5,000 neighbors closest to the set of monosemous and polysemous terms. Both patterns seem to accommodate a Zipf distribution, in which a small percentage of neighbors account for high similarities with each word. In addition to this informal preliminary remark, the results seem to show the plausibility of our hypothesis, where polysemous words have lower similarities to their closest neighbors (see the distribution of ranks-similarities in Figure 3). Beyond 450 neighbors (around cosine .25) this pattern changes, and they tend to have greater similarities to the other neighbors.[2]

To ensure that the curves for the two groups are different, we followed the previously described method. This is meant to show that the curves that describe both kinds of terms display statistically significant differences. Firstly, each term distribution in the set of polysemous and monosemous words is adjusted to Napierian logarithm functions. The average goodness of fit of empirical curves to the Napierian logarithm function .94 for the monosemous group and .97 for the polysemous group, thus justifying use of this function as opposed to others.

Later on, an average difference contrast was performed on the two parameters describing the functions of these curves. Both *a* and *b* for monosemous words are different from those for polysemous words. From the graph of averaged words we expect the slope for the functions of monosemous words to be less steep. This translates into a parameter *a* of monosemous words significantly lower than that of polysemous words. Using the t test to check this unilateral contrast we obtained the statistic ($T(46) = 1{,}91$, $p = 0{,}031$). This also translates into a higher parameter *b* for monosemous words. The t test for this second unilateral hypothesis returns a value of ($T(46) = 1{,}69$, $p = 0{,}049$). These differences confirm what was visually apparent in Figure 4: monosemous words activate their proximal neighbors more strongly but lose this advantage in the proximal set.

In addition to the adjustment to parameters *a* and *b*, and the statistical verification of whether distributions of both word types are different, range distributions allow us to reflect the brute advantage of each word type in each zone. To do so, we use the average distribution of the similarities represented in Figure 4. Both distributions are also adjusted to the Napierian logarithm function, obtaining parameters for each word type.

On the basis of those theoretical functions, the way of calculating brute activations in both zones and the difference in activation by word types is as follows:

Firstly, the cut-off point k between both functions was found. This is done simply by equaling both theoretical functions and calculating the value of *x* at the point where *y* is the same for both (Formula 3).

$$-a \log(x) + b = -a' \log(x) + b' \qquad \text{(Formula 3)}$$

a and *b* being the parameters of the logarithmic function for a word type and *a*' and *b*' being the parameters of the logarithmic function for the other word type.

The solution to this equation is the cut-off point k. This point will delimit the zones in which a function is superior over the other one (also termed the proximal and the distal functions by us). In addition, we have to delimit the upper limit for the distal zone, which we call S. This will be the point where the function with least activation in the distal zone becomes 0. The procedure is equaling this function to zero and identifying at which value of *x*, *y* becomes negative (Formula 4). In this way, the model does not envisage ranges with a negative activation, constraining activation of the first stage to its being excitatory at all times.

$$-a \log(x) + b = 0 \qquad \text{(Formula 4)}$$

With these two values (k and s), we identify the cut-off point for both functions and the higher limit of our analysis. In order to find that extent of the advantage or disadvantage of a word type over another one, we calculate the area between both functions in each zone. It should be borne in mind that in each zone delimited by k, one function will always be above the other one at all times. Thus, in order to find the extent of the advantage, the areas below both functions in each zone were calculated and subtracted. Definite integrals were used to calculate these areas. Each zone has different integration limits. From neighbor 1 to K and from K to S (see Formulas 5 and 6).

$$\text{Proximal Advantage} = \int_1^k (-a \log(x) + b)\, dx - \int_1^k (-a' \log(x) + b')\, dx \qquad \text{(Formula 5)}$$

$$\text{Distal Advantage} = \int_k^S (-a \log(x) + b)\, dx - \int_k^S (-a' \log(x) + b')\, dx \qquad \text{(Formula 6)}$$

The results of this analysis for the monosemous and polysemous word functions are the following: as regards average functions, the best adjustment for monosemous words is $f(x) = -0.085 \log(x) + 0.78$ and for polysemous words it is $f(x) = -0.073 \log(x) + 0.69$. Thus, in the entire previous procedures the values as as follows: $a = 0.085$ and $b = 0.78$. In addition, $a' = 0.073$ and $b' = 0.69$. As regards the cut-off point between both functions, we find that k = 971. After this value, monosemous words lose their advantages and polysemous words generate more activation in their neighbors. As regards the upper limit imposed on the analysis, we find that S = 9433. In this way, we delimit the zones where a function is superior to the other one.

As for the extent of the advantage or disadvantage of monosemous words, the results are as following: in integration limits [1–971] the result is 11.79. The positive sign indicates that this is an advantage for monosemous words, which can also be visually verified in the graphs (Figure 3). When it comes to integration zone [971–9433], things change. Subtraction of its integrals yields a result of −153.38, which is the disadvantage of monosemous activation in the distal zone. This figure is much higher than the advantage of monosemous words in the distal zone, although it is understandable, given that the distal zone is much more numerous. To give a ratio, the area of advantage of polysemous words over monosemous words (distal zone) is 13 times larger than the area of advantage of monosemous words over polysemous words (proximal zone). The first conclusion is that, in the first stage, with no inhibition, the advantage of polysemous words in generation of spurious activation in the distal zone is clearly greater than the advantage of monosemous words in activation of their first neighbors in the proximal area. This may account for the superiority effects in tasks which only require brute activation, with no inhibition mechanisms.

Third Hypothesis: Meaning Hypothesis Generation in the Proximal Set
In the same way as in the two first hypotheses, this time the monosemous word group was compared with the polysemous word group. This time, the cost of generating a meaning hypothesis, measured through the average number of cycles required, the average correlation between the first and last activation vector, and the final activation average were compared.

All this within a framework provided by the C-I algorithm. It should also be remembered that the number *n* of neighbors taken from the proximal neighbor set to constitute the matrices and vectors is subject to two criteria.

Results show the following: Using the rank criterion (Figure 4), results show that there are differences between polysemous and monosemous words. As regards the number of cycles until a stable state is reached, significant differences were found when *n* is 100. Given this condition, polysemous words require a significantly higher number of cycles to reach that stable state (see Figure 4). In addition, the variable standing for correlation between the initial and the final state (State Correlation) was significant when *n* is 300 and 400. Final vectors for ambiguous words are less correlated to their initial activation vectors, and a higher cost is assumed in transition from one to the other, as the number of cycles seemed to suggest. Moreover, the final activation vector for polysemous words has a lower activation mean.

Using the activation criterion (Figure 5), we find that the number of cycles does not differ significantly, nor does the correlation between the initial and the final state. Statistically significant differences are found only in the activation on the final vector: the final vector is more highly activated in the case of polysemous words, which is contradictory with what was previously stated. It is reasonable to suppose that this discrepancy is very probably due merely to the fact that, taking the activation criterion, the number *n* of polysemous words is lower than that of monosemous words, penalizing the final average of the monosemous words. Figure 4 shows that, limiting by activation, monosemous words have more neighbors in the lower activations. We believe that this discordance is merely a methodological issue. In any case, the remaining variables of the activation criterion, although not significant, all point in the same direction of the rank criterion.

Taken globally, there is a clear disadvantage for polysemous words when building a meaning, which ratifies the previous hypotheses, namely that the advantage of monosemous words in the first ranks, less numerous but more significant for meaning generation, may account for the advantage displayed in some experimental studies as regards penalization of polysemous words.

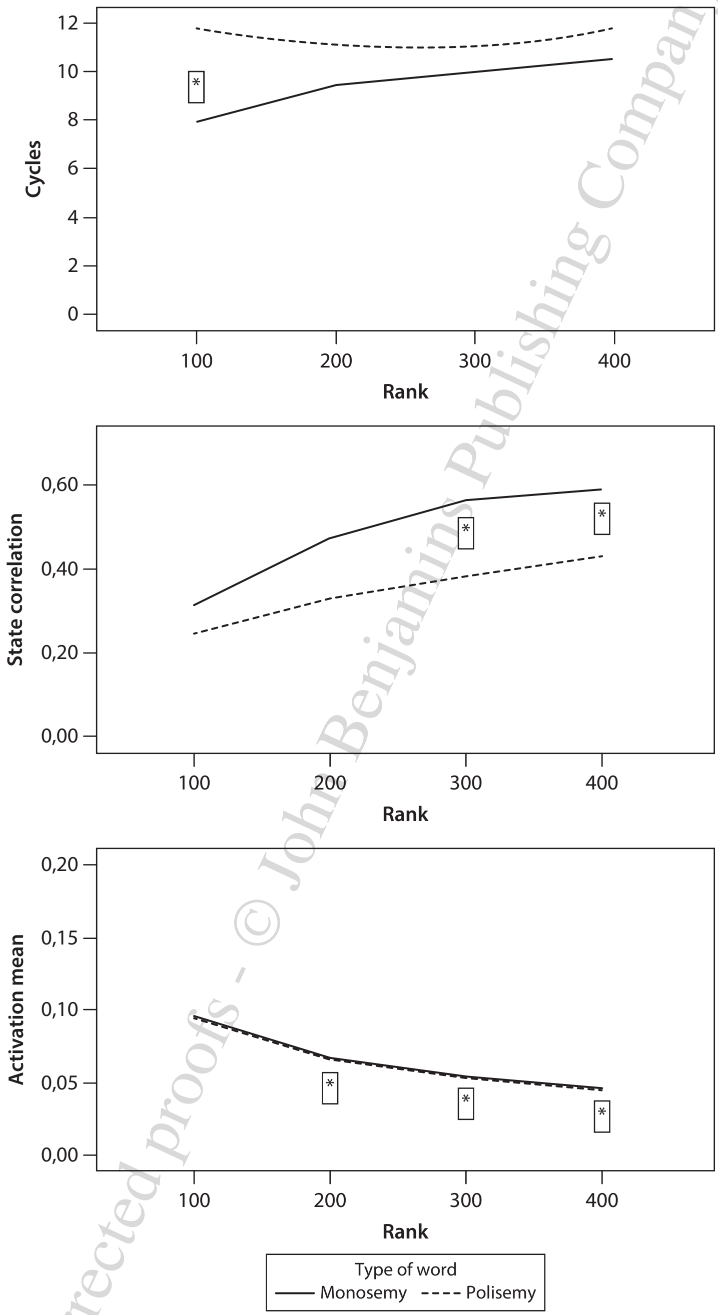

Figure 4. Results of the various variables under the Rank criterion. Horizontal axes show the threshold selected (100, 200, 300 and 400). * means $p <. 005$. Vertical axes show number of cycles, state correlation and activation mean respectively.

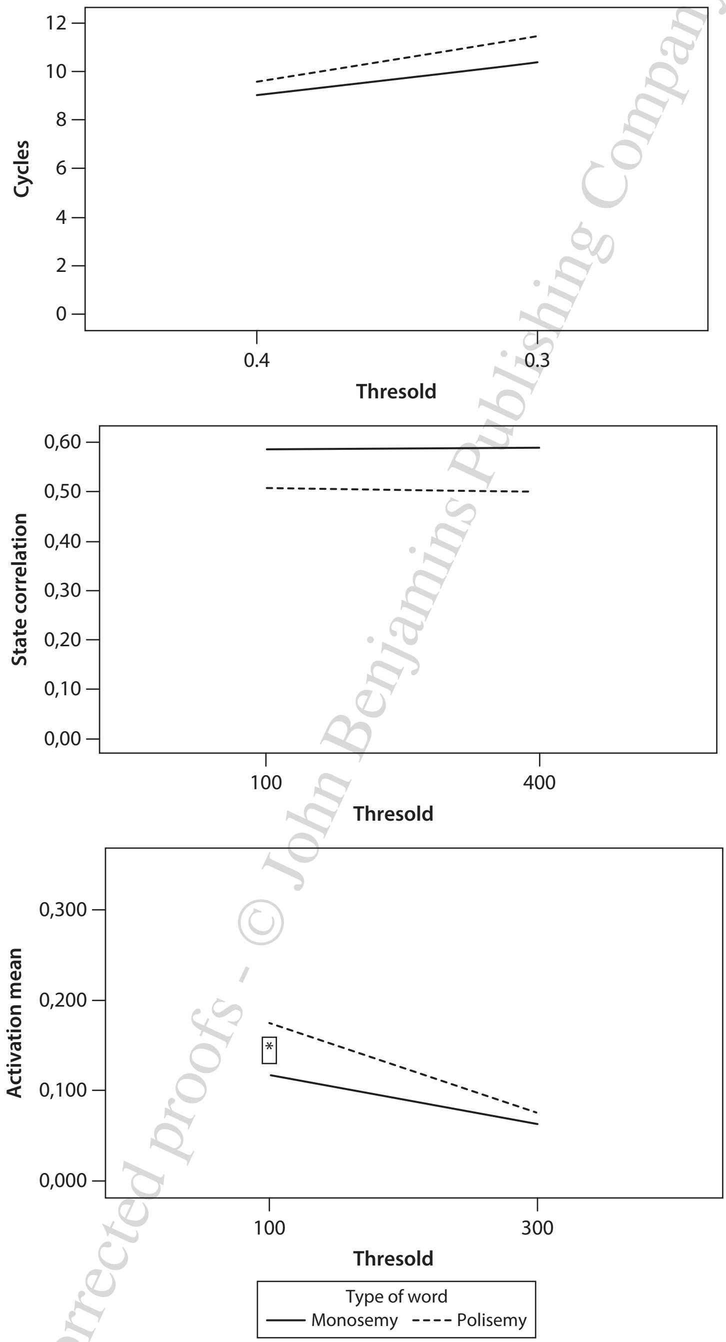

Figure 5. Results of the various variables under the Activation criterion. Horizontal axes show the thresholds that were selected (0.4 and 0.3). * means $p <. 005$.

Discussion

In this first study, we have tried to show by means of a computational model the possible differences in representation which may account for response divergences between monosemous and polysemous words in some experimental data. To this end, the model has been subjected to three analyses. In the first one, we found that polysemous words have a higher degree of entropy, that is to say, they objectively have a lower degree of thematic focalization. This, rather than being a new contribution, supports other experiments, as it is a fact ascertained in other studies which take, for instance, indexes analogous to this one, such as the contextual diversity index (e.g. Adelman et al., 2006). In the second analysis, we found that the early activation distributions were different for each type of word. And in the third analysis we found that there are differences between costs when monosemous and polysemous words are subjected to excitation/inhibition mechanisms. It is in this third analysis that a slight discrepancy appears as regards the criteria used to select the number of neighbors *n* involved in said mutual constriction mechanisms and their theoretical consequences. Therefore, before proceeding to data comparison and explanations, we should examine this point in greater detail.

When the activation criterion is chosen, that is, when neighbors are selected up to a similarity threshold, monosemous words are forced to have a higher number of neighbors. This can be easily seen in Figure 4 (similarity-rank distribution). If a line is drawn in threshold 0.3, it can be seen that monosemous words do not cross the threshold unless it is at a rather higher rank number than polysemous words. This means that a higher number of neighbors will constitute the mutual constriction network. In fact, in the first threshold, the 0.4 threshold, monosemous words have an average of 145 neighbors, as opposed to the 85 neighbors of polysemous words. Likewise, in the second threshold monosemous words have an average of 293 words and polysemous words 235. This is certainly a variable which has an impact when launching activation-inhibition mechanisms. The other alternative was taking the criterion that is usually followed when implementing algorithm C-I (Kintsch, 1998; Kintsch & Patel, 1999; Kintsch, 2000; Kintsch, 2001; Kintsch & Bowles, 2002), which is building the network using a pre-assigned number of closest neighbors. But in this second alternative, activation is also taken into account, as the activation vector for every neighbor is assigned its degree of similarity with the reference term from the start. That is to say, not only is a possible criterion for neighbor selection accounted for, but each one of those neighbors is also already assigned an activation to be subjected to activation-inhibition mechanisms. In our view, the rank criterion is a more natural form, as it keeps the number of neighbors blocked while preserving activation. Actually, if this was applied to all lexicon terms (and not only 5000), there would be no difference in the number of neighbors, but rather in activation of those neighbors.

So in our view the latter criterion is more valid, but we also wanted to use the former one to study its consequences. Given the plausibility with the empirical data of the results obtained using the rank criterion, we predict that in semantic activation-inhibition mechanisms for proximal neighbors the number of neighbors is not as important as distributional properties and the initial activation of neighbors.

Returning to the main results, our first finding when checking the similarity-rank distribution was that, in addition to having a higher degree entropy (Figure 3), ambiguous words have fewer similarities with their first neighbors. This might have consequences for meaning generation, as it is precisely these proximal neighbors that are posited in many models as being involved in activation-inhibition mechanisms (Mirman & Magnuson, 2008). For this reason, and in order to ascertain the consequences, we have created a C-I type network using these neighbors in which this kind of mechanisms would be at work. The results show disadvantages in meaning generation for polysemous words, which incur in higher costs, inferred with the proposed variables: initial state-final state correlation and number of cycles, final activation. As was previously explained, these disadvantages are probably caused not by the higher density of the semantic neighborhood (the number of neighbors) but rather by the distributional properties of each word, which promote less intense relationships with terms with dispersed contents. Probably, due to this dispersal, the number of cycles is higher, as is the cost of moving from an initial to a final state.

Secondly, examining the similarity-rank distribution, we find the inverse pattern beyond a certain threshold — in this case from approximately the 400th neighbor (a cosine of 0.2) onward. The disadvantage of ambiguous words becomes an advantage after an initial set of neighbors (in the distal set), beyond which they activate other words more strongly, causing less specific but much more numerous activations.

These two facts seem to fit well with the phenomena predicted by Piercey & Joordens' (2000) "efficient then inefficient" account: a polysemous advantage in the lexical decision task and an polysemous disadvantage in comprehensive reading. The scenario would be the following:

In a first stage we would have non-specific activation assigned to each term, the same which is displayed in the Similarity-Rank distribution (Figure 4), which shows how all words are activated in a first phase of lexical access. If we assume that this initial, massive and non-inhibitory activation is dominated by distant neighbors (Mirman & Magnunson, 2008), ambiguous words have a measurable advantage in our model, as was shown in the difference between parameters and in the calculation of the areas in the critical zones on the functions that best fitted the model outputs. This can be an advantage for ambiguous words, achieving better response times in LDT (Besner & Joordens, 1995; Joordens & Besner, 1994; Piercey & Joordens, 2000) or else high context distinctiveness tends to lead to longer lexical decision latencies (McDonald & Shillcock, 2001). Lexical decision is conceived as something that precedes full semantic analysis and would only require massive and non-inhibitory activation. In the case of polysemous words, the different activation pattern (represented in the function of each curve of the model) may be enough to break through the threshold that leads to an earlier decision than for monosemous words. In other words, greater similarities with more distant neighbors might mean that polysemous words achieve a higher activation level (more activation from distal neighbors) and break through the familiarity threshold earlier.

However, in a second stage, once the first non-specific activation is launched, inasmuch as some contents are required to be differentially inhibited and activated — contents which are mainly located in the proximal set — the differences between polysemous and monosemous words in that set may justify the costs of achieving a meaning hypothesis. This is shown in the third analysis performed in the proximal set, explaining longer fixation times for polysemous words in reading for comprehension (Duffy et al., 1988). Polysemous words would have a higher cost to reach a meaning hypothesis, as they would be farther from the final differential activation state, that is to say, farther from the final representation. In fact, this proximal set of neighbors is regarded as crucial for the activation-inhibition mechanisms which generate meaning (Kintsch, 1998; Mirman & Magnuson, 2008; Piercey & Joordens, 2000).

Study II: Abstractness

By definition, abstract words differ from concrete words in their amount of perceptual information. Many studies have focused on the availability of mental imagery, but one less explored dimension of the phenomenon examines the occurrence of a given word in a range of contexts: the distribution of each word among contexts or situations. Like polysemous-monosemous words (see the findings of McDonald & Shillecock's, 2001), abstract words may also differ from concrete words in terms of the distributional properties of their representation. The fact that abstract words could have more contextual diversity than concrete words — a concept that is directly proportional to the number of contexts a word appears in (Adelman et al., 2006) — might lead us to revisit the theory, positing that the differences between abstract and concrete words in some behavioral data are partly due to such distribution among possible contexts, as for ambiguous words. In fact, this variable has a demonstrable effect on response times independently of imageability (Adelman et al., 2006). Although McDonald & Shillcock (2001) finally found no correlation between concreteness and their measure of Context Distinctiveness, this was a prediction of the their study: if multiplicity of interpretations/contexts causes a response time advantage for ambiguous words in LDT, why does this not extend to abstract words which also have multiple interpretations/contexts? The answer, however, might be trivial: abstract words differ from concrete words in ways other than number of interpretations/contexts, of which one is perceptual information. But the answer is not so trivial if we take into account that such asseverations (in which the primary representation is the key) have been refuted by some studies. Samson & Pillon (2003) used a dual task, where participants had to retain a square matrix as a means to avoid mental imagery, and found no evidence of the use of imagery in LDT. The size of the concreteness effect did not diminish in the visual interference condition, so there was no evidence that participants relied on visual imagery. Using brain imaging, Pexman, Hargreaves, Edwards, Henry, & Goodyear (2007) also found that when tasks require more sophisticated semantic analysis, there was no difference in the activation of cortical areas, so they suggested that the difference between abstract and concrete words does not lie in their imageability. Some authors even reported years ago a decrease in the use of imagery as children become older (Schwanenflugel & Akin, 1994) and some recent studies have shown the need to introduce task dependency when we talk about use of imagery in linguistic processing (Louwerse and Jeuniaux, 2010).

All of these studies claim that in some cases, abstractness/concreteness could be explained with no reference to imageability, a factor that has been used to distinguish concrete words from abstract words in some of the most classic theories such as dual coding (Paivio, 1986, 1991). These theories state that concrete words have support from primary sensorial representations. For this reason, it is argued that abstract words and concrete words are represented in different networks. Following this conception, some authors have explained the disadvantage of abstract words in activating other words, primarily by semantic similarity (Crutch, Ridha, & Warrington, 2006; Crutch & Warrington, 2005), and the fact that abstract words activate their associated pictures (the non-categorical or non-synonymous relationships) faster than concrete words (Duñabeitia, Avilés, Afonso, Scheepers, & Carreiras, 2009), by means of qualitative differences in the representation, assuming that each kind of word is represented separately in a different net. One network is governed by semantic relationships and the other one by associative relationships.

In contrast, by leaving imageability out of the explanation, we align ourselves with second order isomorphism theories (see Shepard, 1987), which propose that the meaning of a linguistic symbol is not defined in terms of other levels of representation (perceptual representations), but rather by its relationship with other symbols. In other words, verbal information is largely a reflection of perceptual information, and this verbal information is enough for us to process without access to other representational levels such as perceptive factors (Kintsch, 2008). The fact that a purely linguistic model such as LSA, without references to any perceptive aspects, has properly simulated human representation and has been productive in some human tasks such as judging similarity (Landauer & Dumais, 1997) seems to suggest that the theories that propose such an isomorphism in some situations (though not in all cases) are in no way ridiculous. For instance, Louwerse (2009) shows that exclusively linguistic representations encode geographical information for distances between cities, and again, Louwerse (2010) shows that many results of previously published studies finding evidence in favor of embodied cognition can be used to support symbolic cognition with a more parsimonious explanation. For this reason, he claims that embodied representations are partly directly mapped onto linguistic representations as well. Louwerse proposed an integrating model that incorporates an economic principle: symbolic representations, which also encode embodied relations, are "quick and dirty" but provide a good enough way to achieve quick comprehension of real utterances. This representation may become deep (sensory) depending on the language usage situation. Likewise, Riordan & Jones (2011) made comparisons between different techniques: techniques in which perceptual representations such as featured-based ones are found, as they are hand-coded representations; and co-occurrence techniques, techniques which use only one textual source (Riordan & Jones call these distributional techniques). They find redundant information in both representation types, although function, action, and situation relationships prevail in distributional techniques, but not so much perceptive descriptions such as touch, etc. By contrast, feature-based techniques place more stress on material properties, such as the fact of being metallic, for example. At the end of the study it is suggested that once language is acquired, people extract information mainly from linguistic representations, although they may employ sensory representations for information that is not fully redundant, such as perceptive information.

Given this background and the possibilities provided by the previous study, our aim in this second study is to check whether the penalties described by abstract words can also be emulated by the model used in the first study: a vector model with no connection to the sensory-motor world.

Hypothesis

Generally speaking, we use our exclusively symbolic model (it lacks of connection to the sensory-motor) to put into the test the lexical ambiguity component of abstract words. To this end, and taking the previous study as our model, we seek a behavior of abstract words that is analogous to that of ambiguous words. Our hypotheses are the following:

First Hypothesis
In our model, concrete words will have lower entropy (more global weight) than abstract words. This means that they will have higher contextual diversity.

Second Hypothesis
Precisely on the basis of the first hypothesis regarding entropy, we also suspect that in our model, the semantic similarity of the closest neighbors of concrete words is stronger than for abstract words. However, after a neighbor *n* this situation is reversed. We expect the gradients for abstract words to be significantly steeper than those for concrete words, statistically reflecting the previously described activation behavior in polisemous and monosemous words. If this was the case in our model, we would expect greater activation of abstract words in a first stage of lexical access, with an advantage of abstract words in tasks not involving mediation, such as LDT. Given that the empirical data state otherwise, we would have to assume that an exclusively linguistic activation without embodiment would not suffice to account for the advantage of concrete words in LDT tasks.

Third Hypothesis
We also put forward that given the difference in the distributions of both words in our model, abstract words will have higher costs when extracting a meaning hypothesis from the neighbors in the proximal area. If our model behaves in this way, it can account in a simpler way for the disadvantage of abstract words in tasks in which inhibition mechanisms are triggered, as in the case of naming and judgment tasks (Schwanenflugel & Stowe, 1989), as well as the disadvantage as regards their synonymy relationships described in some papers (Crutch & Warrington, 2005; Duñabeitia et al., 2009). In can also do so without appealing to their lack of connection to the sensory-motor world or to qualitative differences in their representations. Generally speaking, the disadvantage of abstract words in these tasks might be explained by their lexical ambiguity component.

To Summarize
If this were the case, penalties for abstract words would not be due to the concrete words' greater degree of reference to the real world, but rather to the ambiguous nature of abstract words in their relationships with other words. We do not deny that concrete words have more reference to primary representations, but rather feel that tasks such as naming and semantic judgments can be resolved primarily by exploiting linguistic distribution predominantly. We are so cautious regarding this issue that we suggest that the cognitive system may exploit sensory representations if our model does produce greater activation by abstract words in distal neighbors.

Procedure

Due to the previous experimental control, in this study, we test two groups of words (abstract and concrete) taken from a previous study (Duñabeitia et al., 2009).

In total there are 39 abstract words and 38 concrete words.[3] The words are controlled for the following variables: frequency and grammatical category (all were nouns). The LSA semantic space from which the neighbors will be extracted is the same as in Study 1. Again, the semantic neighbors were extracted and the distribution of similarities was analyzed.

Method

The method is similar to that carried out with ambiguous words.

First, the entropy was calculated for each group in the raw occurrence matrix. Second, the first 5000 neighbors were extracted and a list of the similarities (cosines) and each ranked neighbor was made. Then the average of each rank for the words from the two conditions was taken — one average for abstract words and one for concrete words in each rank. The result was a distribution like that of the first study, which was also analyzed according to each word curve's fit with the logarithmic function. A mean difference analysis was also carried out on the two parameters that define this function, to verify the difference between the curves of the two word groups. In addition, the same analysis as in study I was carried out to calculate the areas in the proximal and distal zones. Finally, the C-I algorithm was applied to each term of the concrete and abstract lists in the same way as it was applied in study I. This accounted for the costs incurred in the process of generation of a stable meaning hypothesis.

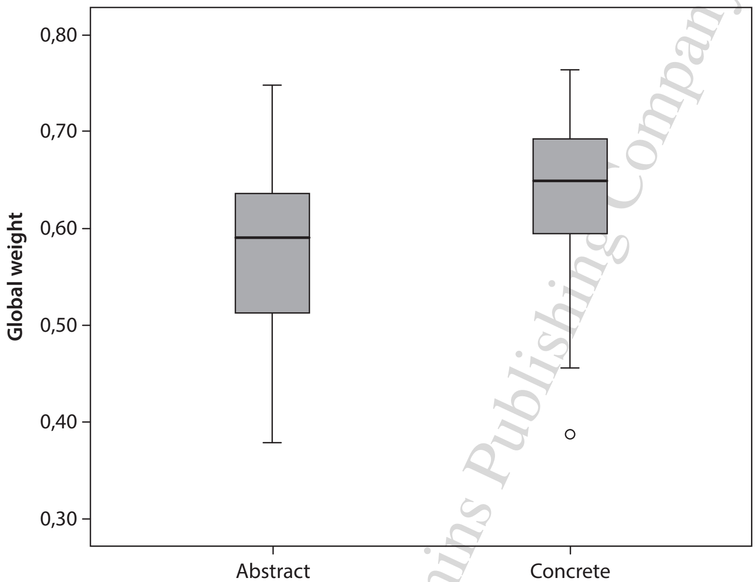

Figure 6. Means global weights for abstract and concrete words. Note that the higher the global weight, the lower the entropy.

Results

First Hypothesis: Global Weight

Our t-test (Figure 6) shows that entropy is greater in abstract words (lower global weight) than in concrete words ($T(75) = 2.42$, $p = .02$). This finding coincides with the earlier results for polysemous words.

Second Hypothesis: Rank Distribution for the First 5000 Neighbors

We find the same pattern as in the first study. Abstract words tend to have lower similarities with their proximal neighbors (Figure 7), with n reaching the threshold where the cosine breaks through .25 (rank 400). After this rank, average similarity for abstract words is higher than for concrete words.

To check whether the curves of the two groups are different, we followed the method outlined in the first study. The goodness of fit for the abstract words group with the logarithmic function was .95, and for the concrete words was also .93, which justifies the use of this function.

Using the above argument and observing the graph of word averages, we expect the slope for functions of concrete words to be less steep. A unilateral hypothesis check was carried out for the parameter *a* of concrete and abstract words, obtaining a significantly higher mean for abstract than concrete words

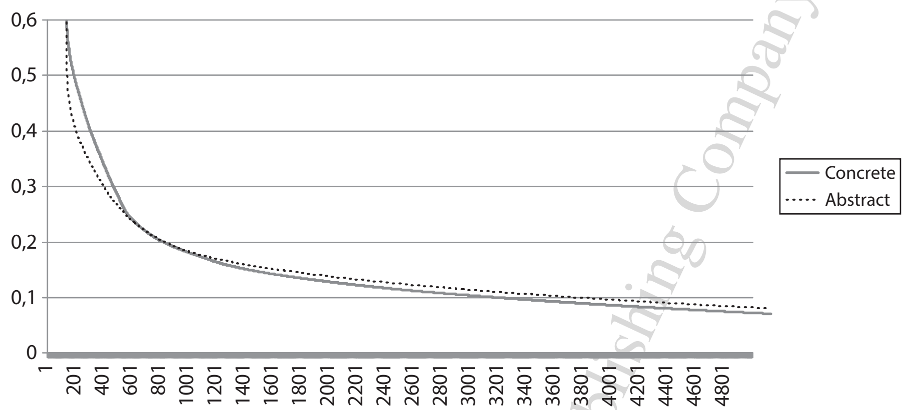

Figure 7. Distributions of ranks and the average cosine, one for abstract words and another for concrete words. The horizontal axis represents the first 5000 ranks (neighbors) sorted by the average cosine of each condition. The Vertical axis represents the cosines.

($T(76) = 1.76$, $p = 0.041$). As for parameter b, the mean obtained was significantly greater for concrete words ($T(76) = 1.69$, $p = 0.047$).

These differences once again corroborate what is visually apparent from Figure 7: concrete words activate their proximal neighbors more strongly but lose this advantage in the proximal set.

As regards calculation of the areas in the proximal and distal zones, respectively, the results are the following: the best fit for concrete words is $f(x) = -0.087 \log(x) + 0.79$ and the best fit for abstract words is $f(x) = -0.072 \log(x) + 0.68$. Calculating values k and s, we find that $k = 1185$ and $s = 8623$. With regard to calculation of the areas, the results are as follows. In integration limits [0–1185] the result is 18.01. The positive sign indicates an advantage for concrete words. When it comes to integration zone [1185–8623], things change. The subtraction of its integrals yields a result of –142.3, which is the disadvantage of activation of concrete words in the distal zone. To give a ratio, the area of advantage of abstract words over concrete words (distal zone) is 7.90 times larger than the area of advantage of concrete words over abstract words (proximal zone). Again, we find the same pattern as in polysemous/monosemous words. In a first stage, without inhibition, the advantage of abstract words in the generation of spurious activation in the distal zone is considerable greater than the advantage of concrete words in the activation of their first neighbors in the distal zone. In this way, activation of the distal zone would provide more brute activation to abstract words.

Third Hypothesis: Meaning Hypothesis Generation in the Proximal Set
The results of the C-I algorithm show that abstract words require significantly more cycles to achieve a stable state than concrete words. These differences are statistically significant all *n* neighbors formed under the rank criterion (*n* = 100, 200, 300 y 400; see Figure 8), and this also happens when the number of neighbors *n* is established taking activation 0.3 as the threshold (Figure 9). This means that abstract words have potentially higher costs when producing a meaning hypothesis. By contrast, no significant differences were found in the other variables, except for average activation, a higher activation of the final vector being achieved in abstract words. This paradoxical effect takes place only if 0.3 is taken under the activation criterion (Figure 9), which reproduces the results of polysemous words under the same criterion. Again, this seems to be a consequence of the criterion chosen, for it should be remembered that a criterion differs from any other in that they use different numbers of words to constitute the C-I network.

Taken globally, these results show a disadvantage of abstract words in the absence of a context when it comes to constituting a meaning, and ratifies the previous hypotheses, namely that the advantage of concrete words in those first ranks, less numerous but more important for meaning generation, may account for the advantage shown in some experimental studies as regards penalties for abstract words.

Discussion

The results show that the penalty of abstract words in terms of relationships with proximal neighbors in a first activation is found in our simulation, which apparently accounts for the disadvantage of abstract words in activating other words, primarily by semantic similarity (Crutch et al., 2006; Crutch & Warrington, 2005). In addition, the inhibitory-excitatory mechanisms launched after the first undifferentiated activation may also partly account for the disadvantage of abstract words in naming and semantic judgments (Schwanenflugel & Stowe, 1989). The mechanism we propose is a simple one: as in the case of polysemous words, abstract words have distributional properties different from those of concrete words, which can be seen in the fact that abstract words have a higher degree of entropy. This has consequences to activate their most proximal neighborhood in the first stage of the process and when launching a differential excitation-inhibition mechanism. The most obvious consequence, which we have simulated in this study, is that abstract words have penalties not only to hold relationships with their first neighbors, but also have a higher cost to generate a meaning hypothesis in later stages. In other words: more resources and more time must be invested

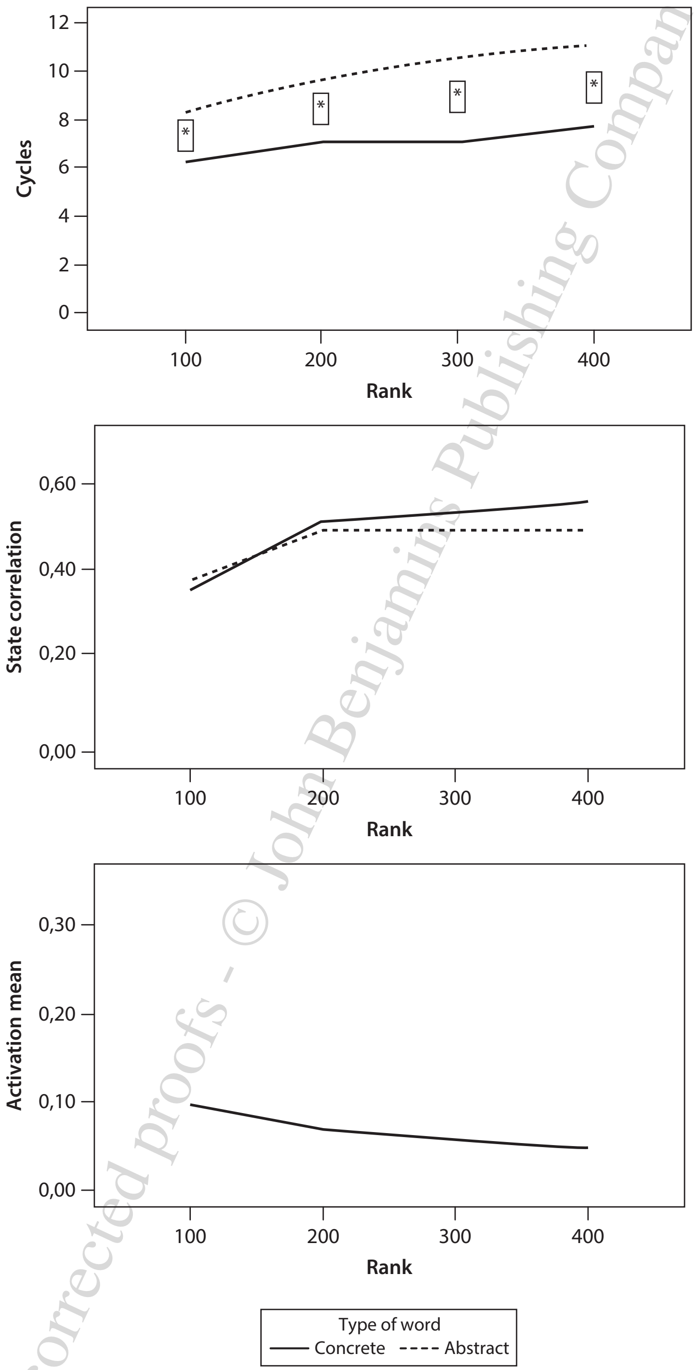

Figure 8. Results for the various variables under the Rank criterion. The horizontal axes show the thresholds selected (100, 200, 300 and 400). * represents meaning $p <. 005$.

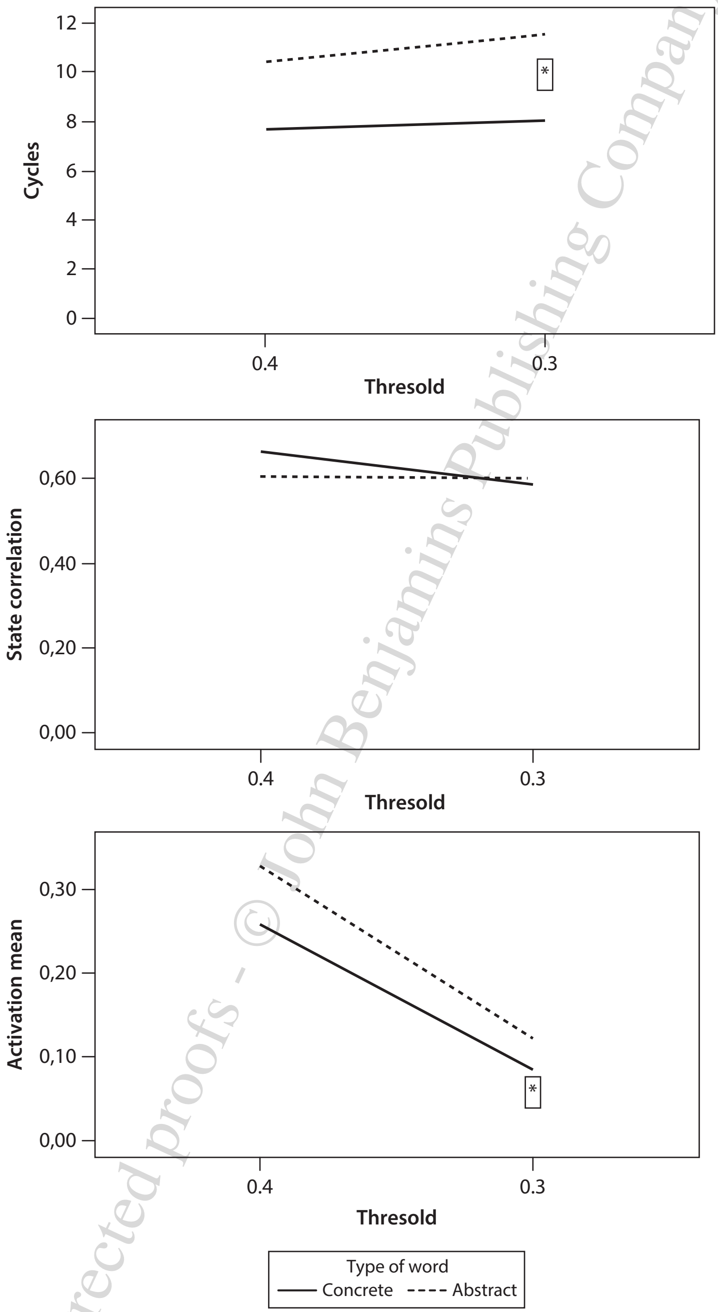

Figure 9. Results for the various variables under the Rank criterion. The horizontal axes show the thresholds selected (0.3 and 0.4). * represents meaning $p < .005$.

to move from an initial to a final state (the meaning hypothesis), as the Context Availability Model (Bransford & McCarrel, 1974) predicts, it being harder to generate contexts and meanings. In fact, in the Schwanenflugel & Stowe study, it is shown that the advantage of concrete words in naming and semantic judgments decreases when abstract words are presented in supportive contexts. This supportive context would also avoid such a penalty in an LSA vector, since a bias toward one meaning outweighs the contextual diversity.

It is also important to highlight that the model seems to reproduce that disadvantage using the same mechanism used with polysemous words, providing us with a parsimonious explanation, with no reach to resort to qualitative differences or sensory representations.

These data fit the models that posit that fast comprehension is possible without appealing to sensory representations (Louwerse, 2011). These models do not deny the existence of connections with the sensory world or that concrete words have certain privileges when they are processed on the basis of this kind of connection. What they suggest is that in certain tasks, greater use is made of this kind of representations, whereas other ones do not make use of them due to a mere question of cognitive economy. Likewise, we do not deny the involvement and integration of sensory representations, but rather intend to make clear that in some circumstances meaning could be accessed in their absence, and perhaps for fast, colloquial adult language it is more effective to do without the grounding in the real world, as adults seem to abandon use of imagery with practice (Schwanenflugel & Akin, 1994). That is to say, that once some childhood linguistic symbols are loaded with sensory meaning, as development progresses, use of the same sensory representations that were used to perform said loading is no longer mandatory, but rather new representations can be used and created without leaving the symbolic plane (Kintsch, 2008).

Therefore, when we say that exploitation of sensory representations is not mandatory in all scenarios, we do not reject their use at all. Indeed, we openly state that our model has no explanation for the fact that concrete words have shorter response times, as has been found in a number of studies, in the LDT task (Bleasdale, 1987; DeGroot, 1989; Howell & Bryden, 1987; Schwanenflugel & Shoben, 1983; Schwanenflugel, Hamishfeger, & Stowe, 1998). According to our model and the curve analysis, abstract and polysemous words would generate a higher degree of activation in the first stage, as they would have the advantage of distal activations, much more numerous and, according to some studies, with no inhibitory capacities (Mirman & Magnuson, 2008). Being rigorous, we find no justification other than when responding to the LDT another source of activation is available, and this is the one arising from sensory representations. More clearly put: even though in naming or judgment use of sensory representations is not necessary, it is necessary in LDF, a task which required imprecise, mass activation in the first processing stage. In fact, some studies seem to suggest that this is not unlikely given that such tasks as naming and judgment do not seem to have correlatives in sensory-motor brain areas, whereas LDT does. Some studies have found that in LDT perceptual areas of the brain are activated, and that this does not happen with other, more sophisticated tasks such as reading for meaning. Binder, Westbury, McKiernan, Possing, & Medler (2005) show that in the lexical decision task, concrete words, and not abstract words, activate areas related to perception. Meanwhile, Pexman et al. (2007) show that when tasks that require more sophisticated semantic analysis are run, there is no difference in terms of the brain areas, but rather the activation of abstract words is broader. These kind of facts are supported also by the findings in Louwerse & Jeuniaux (2010), who find that even though there is both linguistic and perceptual processing, there is a bias in favor of either one of them depending on the type of task or situation, In fact, in an Event Related Potential study, Louwerse & Hutchinson (2012) found that linguistic cortical regions activation precedes perceptual activation in conceptual tasks.

General Discussion

"Efficient then inefficient"

The hypothesis we have tested in the first study is that the vectorial distribution of the terms in an LSA model was enough to account for some particular phenomena concerning ambiguity: on the one hand, that lexical ambiguity (in the absence of a context) promotes processing difficulties when retrieving sense, for example with longer fixation times in reading for comprehension (Duffy et al., 1988) but on the other hand, shorter response time in LDT (Hino & Lupker, 1996; Pexman & Lupker, 1999). This apparent incongruity of results has been formalized in models such as the "efficient then inefficient" account (Piercey & Joordens, 2000). These authors proposed that in order for an advantage for polysemous words to occur in LDT, what happens is simply unspecific activation. The basic explanation was that polysemous words activated more non-specific content than unambiguous in the first step of the process. If what we seek is a response to the lexical decision, polysemous words would have a certain advantage as they already have an non-specific base activation (the non-specific content being much more copious). Conversely, when we seek comprehension in a second step of the process, this non-specific activation should be suppressed and only a small part activated — the part which is most specific to the word itself and its context. It may be that polysemous words need more time to inhibit non-specific activation, as well as more resources to overactivate the most closely-related content which was initially penalized.

To explain all of this formally, we have proposed that on the one hand the contextual distribution of polysemous words penalized the activation of specific content (words that are very closely semantically related) and also, for this reason, polysemous words have a higher cost when generating a solid meaning hypothesis. The contextual distribution of polysemous words is more spread out across their different meanings, as shown by the entropy index, so some meanings penalize associations with words from other senses. On the other hand, our model shows that polysemous words should promote greater non-specific activation, as the penalization for specific content becomes an advantage in associations with non-specific content (less closely related and far more numerous words).

In the first study, carried out with a controlled sample of polysemous and monosemous words, we have simulated these assumptions. It is supposed that in the first stage of access to the lexicon they will receive their corresponding activation. This is simulated by the rank distribution where we find the following: first, the penalization that impedes close relationships with polysemous words makes monosemous words more "efficient" with respect to the proximal set of semantic neighbors. But secondly, we have seen that penalization operates up to a certain number of neighbors. Beyond this, the pattern seems to be inverted and ambiguous words have greater similarity with most remaining neighbors. This second wave of activations, much more numerous than the first one, which according to some studies would not even have inhibitory activity (Mirman & Magnuson, 2008), may allow polysemous words to trigger such non-specific activation, possibly facilitating the strategies adopted during LDT. This seems to fit the studies described above (Hino & Lupker, 1996; Pexman & Lupker, 1999). It is also highly suggestive for some authors who have claimed that deep dyslexia is caused by an inability to inhibit such non-specific activation rather than by difficulties with phonological processing (Colangelo & Buchanan, 2004; Colangelo, Buchanam, & Westbury, 2005). Colangelo & Buchanan argue that this lack of inhibition produces the same effect in deep dyslexics as tasks that require only implicit access to the lexicon in the normal population, such as LDT, especially with ambiguous words.

Moreover, simulating in the proximal neighbor area the excitation-inhibition process which would take place in the second stage of the process, when it is supposed to move from imprecise activation to specific activation, we find that polysemous words have higher costs when generating a meaning hypothesis. In other words, it is assumed that integration processes take place in the first contents of words, what we call the proximal set, and that these are even specifically mobilized in working memory (Kintsch, 1998). In our simulation we found that contents generated by ambiguous words in the form of neighbors have more difficulty integrating and producing an attempted meaning. This seems to fit well the experimental data described.

The discussion on this point makes important assumptions, as lexical ambiguity is almost ubiquitous in human language and it is almost impossible to find an example free from ambiguity. The main conclusion is that the distribution of vector properties itself could explain the advantage of relationships that these might or might not have with other terms, and the way the environment is activated — in other words the way that mentioning a term leads to the activation of specific content, and the form this activation takes. Thus, in contrast to theories that propose different entries for each meaning of an ambiguous word, or even the existence of a common representation and specific representations, we proposed that a model analogous to LSA (a single representation model) might account for some empirical phenomena without casting aside basic assumption: static, context-free entries for terms, with biases based on the moment of acquisition (Kintsch, 2001). They are static in the sense that they emulate permanent knowledge, biased in the sense that the meanings of words are represented in vectors based on their appearance in real language usage, and context-free in the sense that the vectors do not refer to any particular context. More precisely, they are an amalgam of many contexts, and the interaction with one of them is what dynamically generates a meaning (for a review see Jorge-Botana et al., 2010).

Abstract Words

In the second study we also proposed that some of the empirical data concerning abstractness/concreteness may be explained simply using the same assumptions as in the first study. Again, the hypothesis is that the distribution of word properties itself could explain the relationships that an abstract or a concrete term might or might not have with other terms, and the way the environment is activated and inhibited. As with polysemous and monosemous words, abstract words also differ from concrete words in terms of the contextual diversity of their representation, as we have shown in this paper, contrasting the mean of the entropy index for the two groups. For this reason, we have examined whether the empirical data mentioned could be justified using LSA. This is of special interest, as a model like LSA is not only a single representation model for each word, but also makes no reference to the perceptual world. In general, this perceptual world has been postulated as the key difference between concrete and abstracts words, suggesting different networks to represent them.

As regards mass neighbor activation in the first stage (where activation does not yet move toward certain meaning patterns), in the results of our simulation we observed that the similarities between abstract words and their proximal neighbors are less than the similarities for concrete words, emulating abstract words' penalization in activating semantic relationships. In this initial set the abstract words find a penalization, which may justify the experiments where abstract words can facilitate associative rather than semantic relationships (Crutch et al., 2006; Crutch & Warrington, 2005). The consequences of this disadvantage of abstract words in the proximal area are found applying the mutual constriction mechanisms that would take place in the second stage of the process. Once activation is distributed to their distal and proximal neighbors, we have simulated differences if applying an activation-inhibition mechanism in proximal neighbors. The costs of producing a meaning hypothesis seem higher for abstract words, fitting well their disadvantage in semantic judgment tasks (Schwanenflugel & Stowe, 1989).

However, beyond this proximal set of neighbors, this penalization turns to an advantage, allowing abstract words to activate their distal neighbors more strongly. So this is a main discrepancy: our model predicts that abstract words should generate more non-specific activation, and therefore lower latencies for LDT. This seems to not be the case judging by the empirical data, where concrete words seem to have lower response times. The experimental literature reveals the superiority of concrete words in the lexical decision task (LDT) (Bleasdale, 1987; DeGroot, 1989; Howell & Bryden, 1987; Schwanenflugel & Shoben, 1983; Schwanenflugel et al., 1998). One possibility is that concrete words have also a non-linguistic activation — the activation also comes from perceptual representations, and that LDT allusion to this kind of representations is justified. Some studies have found that in LDT perceptual areas of the brain are activated, and that this does not happen with other, more sophisticated tasks such as reading for meaning. Binder, Westbury, McKiernan, Possing, & Medler (2005) show that in the lexical decision task, concrete words, and not abstract words, activate areas related with perception. Meanwhile, Pexman et al. (2007) show that when tasks that require more sophisticated semantic analysis are run, there is no difference in terms of the brain areas, but rather the activation of abstract words is broader. In any case, our speculation regarding the source of activation in LDT requires further confirmation.

To sum up, our approach is close to the theory of the redundancy of perceptual and linguistic information, which posits redundancy of information in both formats, although with a certain degree of specialization in each of them. The key point is that, as people acquire linguistic skills, they extract redundantly-presented information from the linguistic source only, and rely on perception representations only when specific information is needed (Riordan & Jones, 2011). The data in these study suggest that this is possible — that, by extracting redundant information from the linguistic source only, we can account for some (not all) significant effects that are found in the empirical data. Therefore, if this is possible, the second study may support authors who argue that embodied representation is directly mapped onto linguistic representation, and that allusion to primary representations is not mandatory for evoking fast meaning, a meaning that suffices to understand a given situation. Use of supplementary primary representations could be task-dependent (Louwerse & Jeuniaux, 2010) and only be necessary when a specific perceptual comprehension (Louwerse, 2011) is required.

Final Remarks

In summary, this study is primarily a corpus-based study. Bearing in mind that LSA has offered a plausible means of representing lexical units (Landauer & Dumais, 1997; Kintsch & Mangalath, 2011), we have used it in an attempt to represent the penalty or advantage of some types of words in several classic tasks. Since LSA does not respond as a subject would, the conclusions from this approach have their limitations. However, a few analogous behaviors might explain some basic concepts, and help us to understand the place of the symbolic and embodied approaches and distributed models in a future theory of lexical meaning, which is one of the main challenges for a model such as LSA (Kintsch, 2011). We believe that examining the potential responses of a purely symbolic single representation model can show us what may or may not be justified using such a framework. This could be a goal for future LSA research. To extend and contradict a parsimonious model such as LSA might offer a productive starting point for integration of symbolic and embodied approaches in a non-amalgamated manner.

Acknowledgements

The authors wish to thank Javier Valenzuela Manzanares from Universidad de Murcia (Spain) for his useful comments.

Notes

1. They were matched with the rates specified in the LEXESP corpus (Sebastián et al., 2000).

2. The closest neighbors of each word are extracted using the LSA space described above. Each word was compared using the cosine between all words in the space, choosing a list of the 5000 most similar neighbors.

3. It is important to understand that the horizontal axis in Figure 4 (similarity-rank distribution) does not denote temporality: that is to say, a neighbor with rank 1 is not activated before rank 556, but simply potentially receives more initial activation. Therefore, what Figure 4 shows is that activation would be simultaneously received in the first stage each of those 5000 terms, in descending order (higher to lower activation received). As can be guessed, no activation-inhibition mechanism takes part in this distribution yet.

4. 38 concrete words are taken instead of the 39 words in the list given by Duñabeitia et al. as the concrete term "candado" (the noun "lock") was not represented in the semantic space in which we performed the simulations. In any case, we have found that even without that word, the experimental control carried out by them is maintained inasmuch as there are no frequency differences in the LexEsp corpus and as Regards the concreteness index, where both groups differ. These analyses were performed using the BuscaPalabras program (Davis & Perea, 2005).

References

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word naming and lexical decision times. *Psychological Science*, 17, 814–823. DOI: 10.1111/j.1467-9280.2006.01787.x

Besner, D., & Joordens, S. (1995). Wrestling with ambiguity – Further reflections: Reply to Masson and Borowsky (1995) and Rueckl (1995). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 515–519. DOI: 10.1037/0278-7393.21.2.515

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, 17, 905–917. DOI: 10.1162/0898929054021102

Bleasdale, F. A. (1987). Concreteness-dependent associative priming: Separate lexical organization for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 13, 582–594. DOI: 10.1037/0278-7393.13.4.582

Bransford, J. D., & McCarrell, N. S. (1974). A sketch of a cognitive approach to comprehension. In W. Weimer & D. Palermo (Eds.), *Cognition and the symbolic processes* (pp. 189–229). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, W. P., & Ure, D. M. J. (1969). Five rated characteristics of. 650 word association stimuli. *British Journal of Psychology*, 60, 233–249. DOI: 10.1111/j.2044-8295.1969.tb01196.x

Colangelo, A., Buchanan, L., & Westbury, C. (2004). Deep dyslexia and semantic errors: A test of the failure of inhibition hypothesis using a semantic blocking paradigm. *Brain and Cognition*, 54, 232–234. DOI: 10.1016/j.bandc.2004.02.016

Colangelo, A., & Buchanan, L. (2005). Semantic ambiguity and the failure of inhibition hypothesis as an explanation for reading errors in deep dyslexia. *Brain & Cognition*, 57(1), 39–42. DOI: 10.1016/j.bandc.2004.08.018

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain*, 128, 615–627. DOI: 10.1093/brain/awh349

Crutch, S. J., Ridha, B. H., & Warrington, E. K. (2006). The different frameworks underlying abstract and concrete knowledge: Evidence from a bilingual patient with a semantic refractory access dysphasia. *Neurocase*, 12, 151–163. DOI: 10.1080/13554790600598832

Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, 37, 665–671. DOI: 10.3758/BF03192738

Deerwester, S., Dumais, S., Furnas, G., Landauer, T., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41, 391–407. DOI: 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

DeGroot, A. M. B. (1989). Representational aspects of Word imageability and Word frequency as assessed through Word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 824–845. DOI: 10.1037/0278-7393.15.5.824

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27, 429–446. DOI: 10.1016/0749-596X(88)90066-6

Duñabeitia, J. A., Avilés, A., Afonso, O., Scheepers, C., & Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: Evidence from the visual-world paradigm. *Cognition*, 110, 284–292. DOI: 10.1016/j.cognition.2008.11.012

Estévez, A. (1991). Estudio normativo sobre ambigüedad en castellano. *Cognitiva*, 3, 237–271.

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, 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, & Cognition*, 28, 686–713. DOI: 10.1037/0278-7393.28.4.686

Howell, J. R., & Bryden, M. P. (1987). The effects of word orientation and imageability on visual half-field presentations with a lexical decision task. *Neuropsychologia*, 25, 527–538. DOI: 10.1016/0028-3932(87)90077-7

Joordens, S., & Besner, D. (1994). When banking on meaning is not (yet) money in the bank: Explorations in connectionist modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1051–1062. DOI: 10.1037/0278-7393.20.5.1051

Jorge-Botana, G., León, J. A., Olmos, R., & Hassan-Montero, Y. (2010). Visualizing polysemy using LSA and the predication algorithm. *Journal of the American Society for Information Science and Technology*, 61(8), 1706–1724.

Jorge-Botana, G., Olmos, R., & Barroso, A. (2012). The construction-integration framework: A means to disminish bias in LSA-based call routing. *International Journal of Speech Technology*, 15(2), 151–164. DOI: 10.1007/s10772-012-9129-5

Jorge-Botana, G., Olmos, R., & Barroso, A. (2012). Gallito (Version 2.0.1) [NLP Software]. Retrieved from: http://elsemantico.es/gallito20/index-eng.html

Jorge-Botana, G., Olmos, R., & Barroso, A. Gallito 2.0: A natural language processing tool to support research on discourse. In *Proceedings of the twenty-third Annual Meeting of the Society for Text and Discourse*, Valencia from 16 to 18 July 2013.

Kilgarriff, A. (1997). I don't believe in word senses. *Computers and the Humanities*, 31(2), 91–113. DOI: 10.1023/A:1000583911091

Kintsch, W. (1998). The representation of knowledge in minds as machines. *International Journal of Psychology*, 33(6), 411–420. DOI: 10.1080/002075998400169

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psyhonomic Bulletin and Review*, 7, 257–266. DOI: 10.3758/BF03212981

Kintsch, W. (2001). Predication. *Cognitive Science*, 25, 173–202. DOI: 10.1207/s15516709cog2502_1

Kintsch, W., & Bowles, A. (2002). Metaphor comprehension: What makes a metaphor difficult to understand?. *Metaphor and Symbol*, 17, 249–262. DOI: 10.1207/S15327868MS1704_1

Kintsch, W. (2008). Symbol systems and perceptual representations. In M. de Vega, A. M. Glenberg & A. C. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 145–164). Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199217274.003.0008

Kintsch, W., & Mangalath, P. (2011). The construction of meaning. *Topics in Cognitive Science*, 3, 346–370. DOI: 10.1111/j.1756-8765.2010.01107.x

Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language*, 45, 259–282. DOI: 10.1006/jmla.2001.2779

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240. DOI: 10.1037/0033-295X.104.2.211

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, 114, 96–104. DOI: 10.1016/j.cognition.2009.09.002

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, 3, 273–302. DOI: 10.1111/j.1756-8765.2010.01106.x

Louwerse, M. M., & Zwaan, R. A. (2009). Language encodes geographical information. *Cognitive Science*, 33, 51–73. DOI: 10.1111/j.1551-6709.2008.01003.x

Louwerse, M., & Hutchinson, S. (2012). Neurological evidence linguistic processes precedeperceptual simulation in conceptual processing. *Frontiers in Psychology*, 3, 1–11.

McDonald, S., & Shillcock, R. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, 44, 295–323. DOI: 10.1177/00238309010440030101

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(1), 65–79. DOI: 10.1037/0278-7393.34.1.65

Millis, K., & Larson, M. (2008). Applying the construction-integration framework to aesthetic responses to representational artworks. *Discourse Process*, 45, 263–287. DOI: 10.1080/01638530802070015

Nakov, P., Popova, A., & Mateev, P. (2001). Weight functions impact on LSA performance. In *Euro Conference RANLP'2001 (Recent Advances in NLP)*, Tzigov Chark, Bulgaria (pp. 187–193).

Paivio, A. (1986). *Mental representations*. New York: Oxford University Press.

Paivio, A. (1991). *Images in mind: The evolution of a theory*. New York: Harvester-Wheasheaf.

Pexman, P. M., & Lupker, S. J. (1999). The impact of semantic ambiguity on visual word recognition: Do homophone and polysemy effects co-occur? *Canadian Journal of Experimental Psychology*, 53, 323–334. DOI: 10.1037/h0087320

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, 1252–1270. DOI: 10.1037/0278-7393.30.6.1252

Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., & Goodyear, B. G. (2007). Neural correlates of concreteness in semantic categorization. *Journal of Cognitive Neuroscience*, 19, 1407–1419. DOI: 10.1162/jocn.2007.19.8.1407
Piercey, C. D., & Joordens, S. (2000). Turning an advantage into a disadvantage: Ambiguity effects in lexical decision versus reading tasks. *Memory & Cognition*, 28, 657–666. DOI: 10.3758/BF03201255
Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, 3, 303–345. DOI: 10.1111/j.1756-8765.2010.01111.x
Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46, 245–266. DOI: 10.1006/jmla.2001.2810
Samson, D., & Pillon, A. (2003). Concreteness effects in lexical tasks: Access to a mental image? *Brain and Language*, 87, 25–26. DOI: 10.1016/S0093-934X(03)00179-2
Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1998). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, 27, 499–520. DOI: 10.1016/0749-596X(88)90022-8
Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychololgy, Learning and Memory and Cognition*, 9, 82–102. DOI: 10.1037/0278-7393.9.1.82
Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words. *Reading Research Quartely*, 24, 114–126. DOI: 10.2307/748013
Schwanenflugel, P. J., & Akin, C. E. (1994). Developmental trends in lexical decisions for abstract and concrete words. *Reading Research Quarterly*, 29, 251–263. DOI: 10.2307/747876
Sebastián, N., Martí, M. A., Carreiras, M. F., & Cuetos, F. (2000). *LEXESP, léxico informatizado del Español*. Barcelona: Ediciones de la Universitat de Barcelona.
Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237, 1317–1323. DOI: 10.1126/science.3629243
Wandmacher, T. 2005. How semantic is latent semantic analysis? In *Proceedings of RECITAL'05*, Dourdan, France.

Appendix

Monosemous words: lunes, comedia, nieve, paro, acera, sonido, barro, médico, presidente, alumno, cariño, apellido, acuerdo, barba, cultura, broma, árbol, ceremonia, combate, dolor, cansancio, color, animal, abeja.

Polysemous words: planta, hoja, copa, banco, rosa, programa, caja, papel, cadena, bomba, bolsa, diente, media, metro, columna, fuente, división, corriente, corredor, capital, cabo, banda, artículo, lila.

Abstract words: pedazo, ligereza, difusión, creación, artículo, comodidad, detalle, partida, libra, herramienta, idea, carga, elegancia, pecado, anterior, temporal, piso, sirena, llamada, punta, polo, consulta, relieve, cauce, lomo, pasajero, tocar, rabia, infierno, olor, geografía, abrir, luminosidad, hundimiento, velocidad, cera, conductor, rapaz, capullo.

Concrete words: bigote, ciego, compás, triángulo, músculo, muñeca, mármol, nadar, pólvora, ratón, estuche, navaja, taburete, bota, fútbol, cuna, correo, caracol, cascabel, gallina, puño, marido, espina, chaleco, infarto, marfil, tapón, piano, nido, botón, cepillo, cinturón, cenicero, mantel, jinete, puerta, lechuga, pierna.

Corresponding Address

Guillermo Jorge-Botana
Universidad Nacional de Educación a Distancia (UNED)
Departamento de Psicología Evolutiva y de la Educación. Facultad de Psicología
Calle Juan del Rosal, 10 (Ciudad Universitaria)
Madrid, Spain

gdejorge@psi.uned.e