The Semantic Network of the Spanish Dictionary During the Last Century: Structural Stability and Resilience

Camilo Garrido[1,2], Claudio Gutierrez[1,2], Guillermo Soto[3]

[1] Department of Computer Science, Universidad de Chile

[2] Millennium Institute of Foundational Research on Data

[3] Department of Linguistics, Universidad de Chile

E-mail: cgarrido@dcc.uchile.cl, cgutierr@dcc.uchile.cl, gsoto@uchile.cl

Abstract

The semantic network of a dictionary is a mathematical structure that represents relationships among words of a language. In this work, we study the evolution of the semantic network of the Spanish dictionary during the last century, beginning in 1925 until 2014. We analysed the permanence and changes of its structural properties, such as size of components, average shortest path length, and degree distribution. We found that global structural properties of the Spanish dictionary network are remarkably stable. In fact, if we remove all the labels from the network, networks from different editions of the Spanish dictionary are practically indistinguishable. On the other hand, local properties change over the years offering insights about the evolution of lexicon. For instance, the neighbourhood of a single word or the shared neighbourhood between a pair of words. This paper presents preliminary evidence that dictionary networks are an interesting language tool and good proxies to study semantic clouds of words and their evolution in a given language.

Keywords: semantic networks; dictionary networks; Spanish language

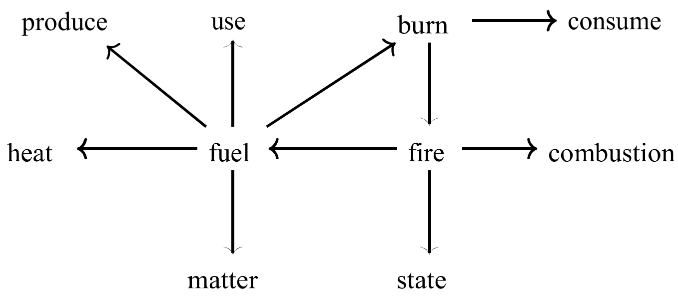
1. Introduction

The lexicon of a language can be organized as a semantic network by considering the words as nodes and the similarities of some kind among the words as representing edges. A suitable proxy to such a network is the one obtained from a dictionary, built as follows: The nodes are the dictionary entries (properly cleaned), and for each entry define an edge from it to all the words that occur in its definition (which, when properly cleaned, occur as entries too) (see Figure 1). These *dictionary networks* are well known and have attracted linguistic interest (cf. Picard et al., 2009; Levary et al., 2012).

Until now the studies of dictionary networks have focused on static versions of dictionaries. But a dictionary evolves over time. New words are added to the lexicon, due to the introduction of a new, previously incommunicable concept, or to increase the different ways of mentioning an existing concept. Additionally, some words experience some slight changes in their meanings to adapt to new cultural trends. A few words are eliminated. Some new organizing criteria are introduced.

The evolution of a dictionary suggests studying the corresponding evolution of its associated network. Are there observable patterns in such evolution that can be of linguistic interest? What can the evolution of such networks tell us about the evolution of the lexicon of a language? These are the types of questions that motivated this research.

In this paper we study the historical evolution along the last century of the networks associated with the most traditional Spanish dictionary. This dictionary has been issued by the Spanish Royal Academy since 1780, with regular periodicity and a rather stable philosophy and methodology.



Fire: Fuel in a state of combustion.
Fuel: Any matter used to produce heat by burning.
Burn: To consume with fire.

Figure 1: The network built from the entries fire, fuel, burn, and their definitions.

We investigate the permanence and changes of structural properties of the network of this Spanish dictionary beginning in 1925. There are two groups of network properties that we explore: global and local properties. The *global properties* are those capturing aspects as a whole and give an overall view of the network, for example, ratio of number of nodes versus edges, connectivity, centrality, etc. The *local properties* correspond to those topological properties of vicinities of nodes, such as clustering coefficient, the number of triangles in a particular location or the similarities and differences between the cloud surrounding two words in dictionary networks.

We highlight two main findings of our study. First, the structural properties of these networks are remarkably stable. Simply put: if we delete the labels of the nodes (i.e. of the words), and normalize the size of the networks, it would be very difficult, if not impossible, to tell which network corresponds to which year. The 1925 and the 2014 dictionary networks have almost the same structure. In particular, these networks are highly resilient, that is, they keep their structure in spite of the deletion of words and local perturbations. Second, the (historically) successive networks offer insights on how the semantic neighbourhood of a word evolves, that is, how relationships among words evolve. As we considered the dictionary as a suitable proxy of the lexicon of the language itself, this could shed light on the evolution over time of particular meanings and senses of concepts. One example we present is that of the noun *sex* (sexo) and adjective *sexual*. Early in the 20th century the word *sex* was in a cloud of biological terms and almost disjoint from *sexual*, which refereed to human behaviours, the former with higher presence than *sexual*. In 2014, the cloud around word *sexual* became bigger than that of *sex* and more words directly connecting both entries appeared.

The paper is organized as follows: Section 2 presents an overview of dictionary networks. Sections 3 and 4 present the structural stability of the Spanish Dictionary network and the changes in local features. Section 5 analyses related work. Section 6 then presents our conclusions.

2. Dictionary networks: an overview

The definition of a word involves recursively new words, senses and meanings. Litkowski (1978) observed that this relation naturally forms a network that has linguistic interest. See Section 5, Related Work, for a more detailed overview of the developments of these type of networks.

2.1 Basic network model

In this work we utilize a simple (naive) model of a dictionary network that lacks any information on the type of word on nodes and edges, that is, just words pointing to other words represented in a standard form. At first sight, this simplification might seem impractical since it misses a lot of linguistic information (e.g. type, morphology, inflection, etc.) present in a dictionary. Nevertheless, several studies have shown the power of this simple model (Clark, 2003; Picard et al., 2009; Levary et al., 2012). In fact, besides facilitating the analysis of the network and its comparison with those in other fields, it captures the main features of the structure of these networks.

For this work we implemented the following procedure to build the networks:

1. *Model or design:* Consider all types of words as a single type: forget if they were nouns, verbs, adverbs, etc. Merge the entries that correspond to the same word into one definition, e.g. *Singer: A machine for sewing cloth.* and *Singer: One who, or that which, sings.* Forget the role and place of occurrence of a word, as well as its number of occurrences, inside a sentence (i.e. transform the defining text of a word in a set of words).

2. *Clean:* Remove entries that are inflected forms, e.g., *singing: from Sing.* Remove prepositions, conjunctions, and articles from entries and definitions. We consider them stopwords. They appear too often in any text and they would add noise to the graph. Lemmatize each word occurring in the definitions (transform nouns into singular; verbs into the infinitive; adjectives into their male singular form). In this work, we used Freeling (Padró & Stanilovsky, 2012) for the lemmatization of Spanish words and StanfordNLP (Qi et al., 2018) for the lemmatization of English words. Finally, remove any word that does not appear in the dictionary, *e.g.* prefixes and suffixes like *Ex-* and *-able*.

3. *Mathematical model of the dictionary:* Build the graph over the previous data. At this point, the dictionary D has become a universe of words W and a set of pairs (w,def(w)), where $w \in W$ is an entry in D and def(w) $\subseteq W$ is the set of words occurring in the definition of w.

4. *Build the network:* From the data in (3), construct a directed graph $G = (V,E)$, where the nodes are $V = \{w/(w,S) \in D\}$ and the edges $E = \{(w,w^0)/(w,S) \in D$ and $w^0 \in$ def$(w)\}$. For example, from the entry "*Eaglet (n.) A young eagle, or a diminutive eagle.*" we get the edges *(Eaglet, young)*, *(Eaglet, eagle)* and *(Eaglet, diminutive)*.

2.2 Main structural features

The network of a dictionary allows one to explore and study the global and topological properties that emerge from the network of words that cannot be captured locally (e.g. considering only isolated entries and their definitions). A classic global property is component analysis that allows finding subgroups of words according to connectivity. It shows four categories (Figure 2):

Giant Strongly Connected Component (SCC), this refers to words that recursively use themselves, which amount to about 1/3 of all words, most of them corresponding to entries never used in a definition. Bidirectional Component, words that mutually use each other in their definitions. Bidirectional Strongly Connected Component, words that mutually use each other and recursively use all other words in the category – amounting to 10% of all dictionary words. And Triangle Strongly Connected Component, triples of words that mutually need each other and recursively use all other words in the category. We will see that these components are stable parts of a dictionary.

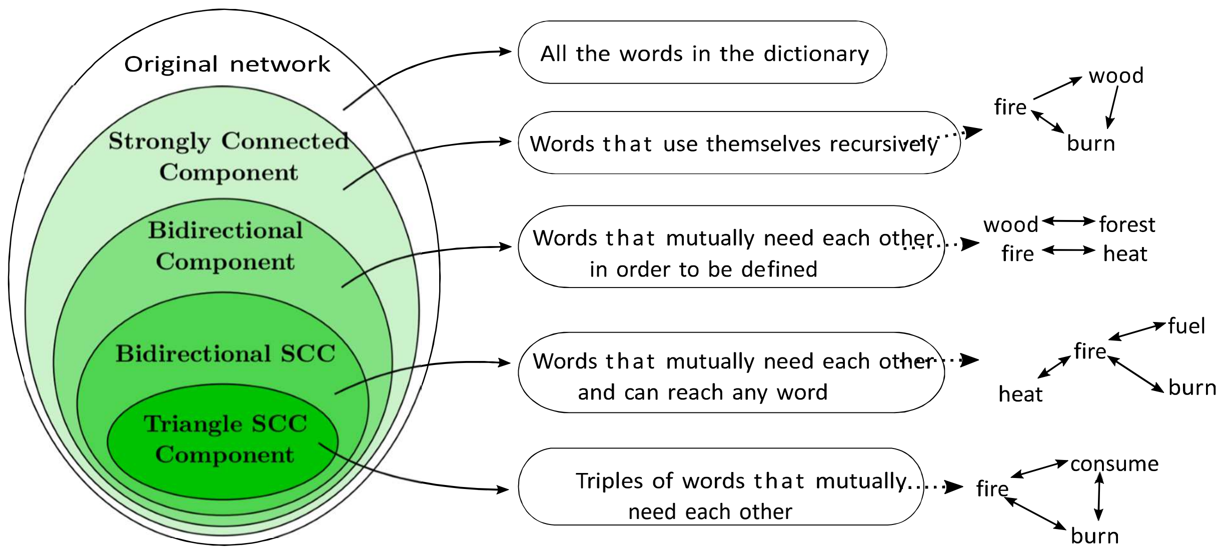


Figure 2: Structural components of a dictionary network. Examples on the right taken from the *Online Plain Text English Dictionary* (OPTED).

3. The Spanish dictionary network: a stable and resilient structure

The Spanish Language dictionary (*Diccionario de la Lengua Española*, DLE) is a dictionary issued periodically since 1780 by the Spanish Royal Academy (currently in its 23rd edition). The new versions present updated lexicon and linguistic and editorial reorganizations[1].

In this section we study the network of the DLE and show that its basic structure remains stable and resilient over the years. We analyze three editions of the DLE: the 15th (published in 1925), the 18th (1956), and the current, 23rd edition (2014). According to the Royal Spanish Academy, the 1925 and 2014 editions are especially significant. The former (1925) incorporates attention to different Spanish-speaking territories besides Spain, and describes simpler definitions. The latter (2014), the most recent version, besides updating the lexicon, modifies its structure to facilitate searches, and incorporates other features, e.g. showing variations of entries and a consistent treatment of their male and female forms. To have an intermediate reference point, with a logarithmic interval between the extremes (30 and 60 years), we employed the 18th edition (1956). We used printed versions (none exist for the 1925 and 1956 editions) and for reasons of space we avoid the description of the tedious work and lessons obtained from scanning, cleaning and tuning the final texts.

3.1 Basic measures

A first snapshot of the evolution of dictionary networks is given by basic network measures (see Table 1) (Newman, 2003). The number of nodes (n) indicates the number of words in the dictionary. The dictionary grows about 15% every 30 years in this period. Edges (m) do not grow at the same rate, and the current dictionary has on average less edges per node (z) than previous years (meaning shorter definitions on average). Despite the changes in the number of nodes and edges, the average distance between entries (l) is not affected, staying around 4. The parameter α, the exponent of the degree distribution function ($p_k \sim k^{-\alpha}$), also remains almost unaffected over the years with the value $\alpha \approx 2.6$. The clustering coefficients over the years are also very similar, both global (c^1) and local (c^2). In dictionary networks, two entries having a common (non-frequent) word in their definitions are likely to be related. Lastly, the degree correlation coefficient (r) indicates whether the high-degree vertices in the network associate link preferentially with other high-degree vertices or not ($r = 1$ means high and $r = -1$ means low connectivity). This coefficient falls over the years. This may be caused by lexicographic decisions between editions, e.g. the removal of adverbs with the suffix *-mente* or past participles of verbs.

[1] http://www.rae.es/diccionario-de-la-lengua-espanola/presentacion

	n	m	z	l	α	c^1	c^2	r
DLE 1925	60,823	1,058,012	17.39	4.03	2.59	0.019	0.227	0.042
DLE 1956	69,719	1,174,912	17.49	4.03	2.58	0.017	0.225	0.039
DLE 2014	87,255	1,076,377	12.34	4.09	2.65	0.015	0.224	0.002
OPTED	95,095	979,523	20.60	4.64	3.13	0.009	0.217	-0.008
WordNet	84,967	1,134,957	26.72	2.99	2.99	0.029	0.203	-0.016

Table 1: Basic measures for the networks of the Spanish dictionary (DLE) over the years. The Online English dictionary OPTED and WordNet networks are shown for comparison. n and m are the number of nodes and edges, respectively; the other parameters are explained in Section 3.1.

3.2 Component analysis

Components are classic features when describing the topology of networks (Section 2.2). For the Spanish dictionary network (Table 2), the Giant Strongly Connected Component for all three editions remains around 30% of the whole network. The Bidirectional Component stays around 17% of all the words over the years. The Bidirectional Strongly Connected Component covers about 11% of the network. Finally, one of the strongest notions of connectivity is the subgraph induced by the strongly connected component of triangles. It represents less than the 3% of the network in each dictionary. The ratio of the size of each component is consistent over time. The words composing the components are also very consistent. Note that around 80% of the words in a component in 1925 remain in the same component in 2014 (Table 3).

3.3 Centrality measures

We tested four classic centrality measures for each DLE network: Betweenness Centrality, Closeness Centrality, Degree, and PageRank (Boldi & Vigna, 2014). The ranks are very similar if we just take into the account the top nodes/words. Here we present the recurrent words (RW) in the top 20 ranking for each measure:

Betweenness (9 RW): *acción, cosa, dar, estar, hacer, mano, parte, ser, tener.*

Closeness (10 RW): *alguno, cosa, dar, decir, estar, hacer, otro, persona, ser, tener.*

Degree (12 RW): *acción, alguno, cosa, dar, decir, estar, hacer, otro, parte, persona, ser, tener.*

PageRank (13 RW): *acción, alguno, cosa, dar, decir, efecto, estar, hacer, otro, persona, poder, ser, tener.*

Over the last century, half of the words stayed in the top 20 ranking. These are basic words that help to put together definitions and the dictionary, e.g. *Natación:* ***acción*** *y* ***efecto*** *de nadar* (Swimming: action or effect of swim).

3.4 Cliques

Cliques are sets of nodes such that any pair among them is connected by an edge. In the context of dictionary networks, cliques are a local property that allows identification of a strong dependency among words (each one occurs in the definition of all others). For example, *cosa, dar, decir, hacer, ser, tener, todo.*

In the Spanish dictionary network, the number of cliques grows from edition to edition, but the growth rate seems to slow down over the years (Figure 4). There are no bigger cliques than K_7 in any of the three editions.

3.5 Resilience of the dictionary network

Resilience refers to the vulnerability or the ability of a network to resist link or node failures. This happens to be a relevant property in dictionary networks. As a notion of resilience, we use the variation of the size of the largest component as nodes are removed from the network. We use two approaches to node removal: random choice and high in-degree nodes, the latter meaning the removal of words that occur the most in other definitions. As baseline, we compare the behaviour of dictionary networks with that of a random graph. We use the random graph model proposed by Barabási and Albert (1999) based on the idea of preferential attachment. It is frequently used for language networks comparison (Dorogovtsev & Mendes, 2001; Steyvers & Tenenbaum, 2005).

	1925	**1956**	**2014**
Original network	60,823	69,719	87,255
SCC	18,307	21,538	26,989
Bidirectional Component	10,462	12,061	16,025
Bidirectional SCC	6,125	7,429	11,308
Triangle SCC	1,033	1,318	2,359

Table 2: Component sizes of the Spanish dictionary networks in number of words.

	1925-2014	**% of 1925**
Original	54,235	89.1%
SCC	15,514	84.7%
Bidirectional Component	7,665	73.3%
Bidirectional SCC	4,841	79.0%
Triangle SCC	828	80.2%

Table 3: Number of words (and percentage) from 1925 that remain in the same component in 2014.

It turns out that removing random nodes produces almost no damage at all. All three dictionaries and random graphs resist the attacks well. The size of the component decreases linearly with respect to the number of nodes removed. On the other hand, dictionary networks and random graphs behave very differently when removing high in-degree nodes.

Dictionary networks resist more attacks than random graphs (Figure 3). Random graphs decline quickly. Removing just 10% of the high in-degree nodes is necessary to completely destroy and scatter the graph. That is not the case with dictionary networks. The giant component of dictionary networks decreases almost linearly until we remove about a third of the network. From that point forward, the giant component starts to decline rapidly, scattering completely when 37% of the high in-degree nodes are removed. It is important to note that the resilience of connectivity of dictionary networks does not rely on frequently used words that connect the network, but on the high connectivity among all words. One could express this by saying that it is very difficult to completely remove a cloud of close concepts; there will always remain other ways to express them. This seems to be a particular property of dictionary networks, as results for other real world networks do not show this behaviour (Jeong et al., 2001; Dunne et al., 2002; Newman et al., 2002).

4. What changes: the local features

Despite its structural stability, there are changes in the successive versions of the DLE: new entries are incorporated, some entries are removed and some definitions are enriched or modified. In this section, we focus on these changes in the dictionary.

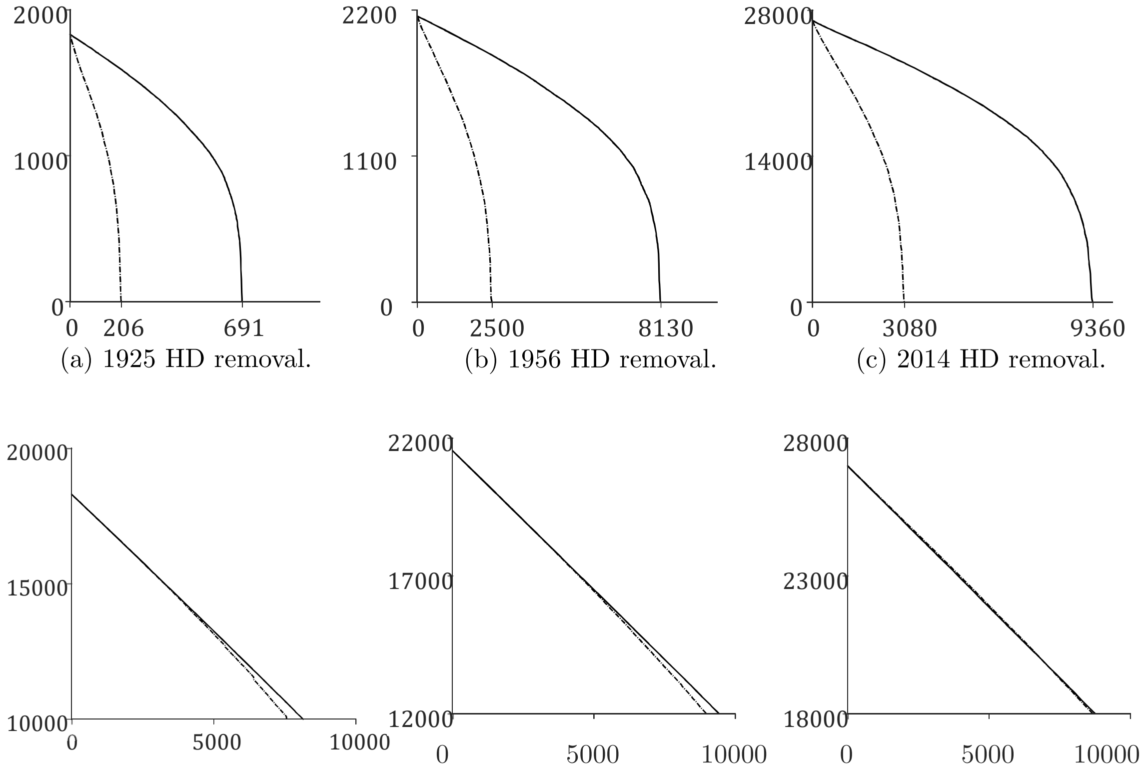
4.1 Definitional and interchangeable entries

The entries in the DLE can be divided into two groups: *definitional* entries are words used to define other words and *interchangeable* entries correspond to words that do not occur in any definition at all. In network terms, definitional words are those that have inlinks and outlinks, while interchangeable words have only outlinks. The fact that a word has only outlinks means that in some sense is "disposable", that is, it could be replaced by the words in its definition (Levary et al., 2012), hence the name interchangeable.

If we study how incorporations and deletions of entries from one version of the dictionary to another occur, eight possible outcomes show up (Figure 4). Definitional entries can (1) stay as a definitional entry, (2) become an interchangeable entry, (3) be removed from the dictionary. Likewise, interchangeable entries can (4) stay as an interchangeable entry, (5) become a definitional entry, or (6) be removed from the dictionary. Additionally, new entries are incorporated into the dictionary as (7) new definitional entries or (8) new interchangeable entries.

	1925	1956	2014
K_3	2,208	3,007	5,911
K_4	489	917	1,311
K_5	95	299	347
K_6	10	69	69
K_7	1	8	5

Table 4: Cliques in DLE networks.



(a) 1925 HD removal. (b) 1956 HD removal. (c) 2014 HD removal.

(d) 1925 random removal. (e) 1956 random removal. (f) 2014 random removal.

Figure 3: Sizes of the giant component as nodes are removed. On the left, high degree (HD) node removal. DLE network (solid line) keeps its structure (giant component) as compared to a random network (dotted line). On the right, random removal does not affect the size of the giant component in either DLE or a random network.

Most of the entries in a dictionary do not change their type between versions. In fact, in the DLE (with new versions approximately every 30 years) between 80%-90% of definitional entries stay as definitional, and a similar percentage of interchangeable entries stay as interchangeable (1 and 4 in Figure 4). When new words are added to the dictionary, most of them (76%-95%) enter as interchangeable (8 in Figure 4); only a few of them occur in definitions (7 in Figure 4). On the other hand, almost all of the entries that are removed from the dictionary were interchangeable entries (6 in Figure 4).

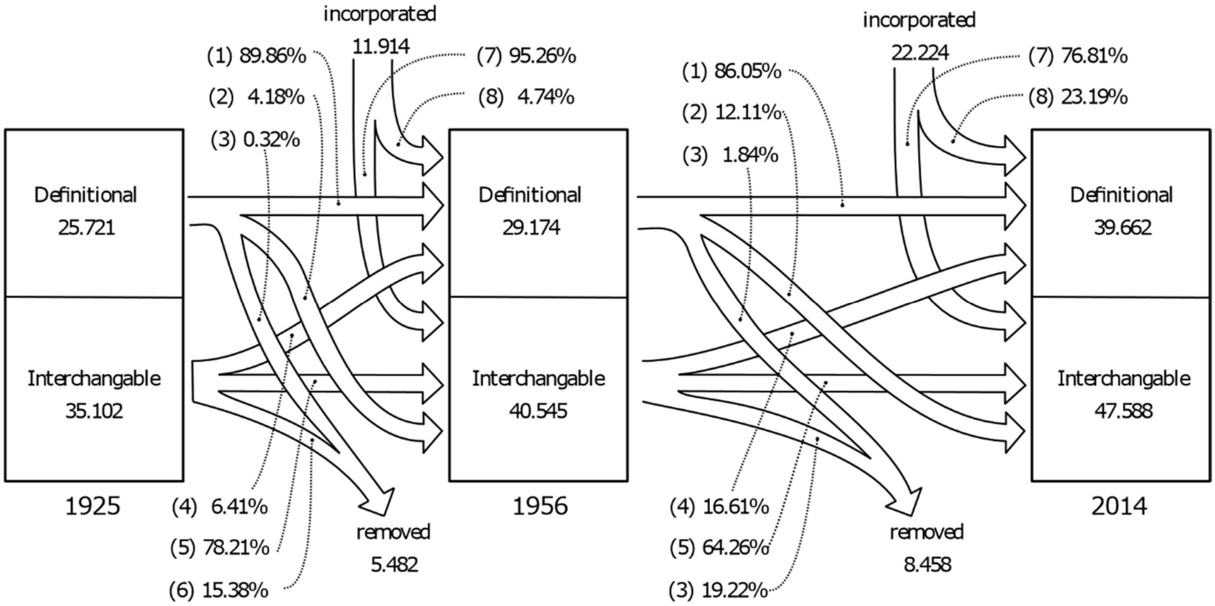


Figure 4: Changes in the entries of the DLE from 1925 to 1956 and 1956 to 2014. For a detailed explanation of the figure see Section 4.1.

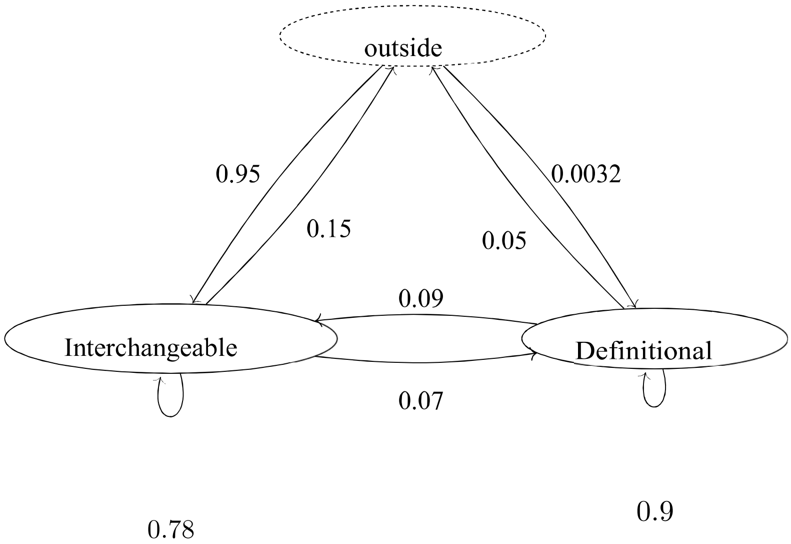


Figure 5: Markov chain that describes the probability of transitions among types of words every 30 years in the Spanish Dictionary.

In order to better describe the transitions among the types of words, we build a Markov chain using the empirical data of the transitions over the years (see Figure 5). A Markov chain is a stochastic model that describes the transitions between possible states using only its current state. It can be described as a directed graph with probabilities for edges and states for nodes. A word can be in one of three states. It can be a definitional, it can be interchangeable, or it can be "outside". The state outside means that the word is not in the dictionary. This model allows us to estimate the probability of a word being in a state in future editions of the Spanish dictionary and the paths it is going to take. For example, a definitional word has a probability $p = 0.9$ of staying as definitional in 30 years in the future (one iteration). If we consider a span of 90 years (three iterations), a definitional word has a probability of $p = 0.729$ (calculated as $0.9 \cdot 0.9 \cdot 0.9$) of always staying as definitional. The model allows us to calculate the probability of more complex transitions. For example, the probability of a definitional word becoming interchangeable in one iteration and then being removed from the dictionary in the next iteration is $p = 0.0135$ (calculated as $0.07 \cdot 0.15$).

4.2 Examples of simple local changes

These changes do not affect or change the overall structure of the network (as we saw in Section 3). But they impact at the local level. In fact, these changes alter the structure of the vicinity of some words (not only those whose definition explicitly changes). We will illustrate these changes through some examples in order to offer insights on how the evolution of the network structure speaks about semantic features.

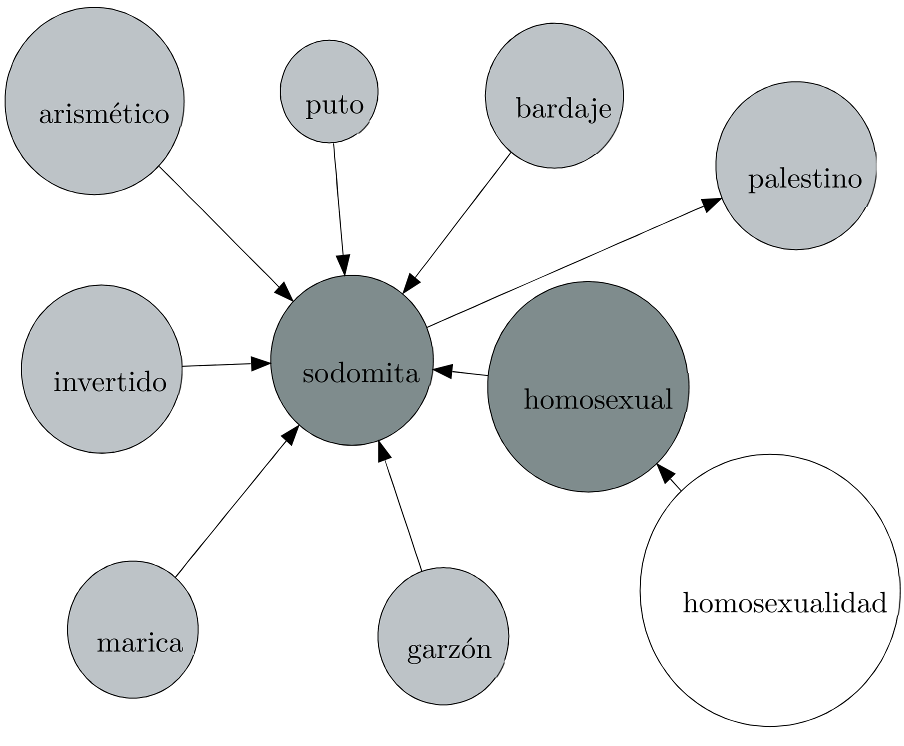
First, entering and outgoing words. *Aeropuerto* (airport) is an obvious case of an entering word that was not present in the 1925 edition. In fact, airplanes and other aerial words were emerging concepts at the time. In 1956, *aeropuerto* is already incorporated as a definitional entry. Later, in 2014, *aeropuerto* is still a definitional entry being used by 17 different words in their definitions, such as airfield (*aeródromo*), checkroom (*consigna*), and tower (*tower*). On the other hand, there are words that were slowly put aside in the dictionary. These words were definitional entries in 1925. In 1956, they became interchangeable entries, as they did not appear in any definition. And in 2014, they were completely removed from the dictionary. Examples are *Adolecente* (old form of adolescent); *fecundante* (someone who impregnates or fertilizes); *escaza* (an Aragonese word refering to a certain type of pot).

Second, words whose cloud of meaning changes. Consider the word *prostituta* (prostitute). The 1925 dictionary contains the definitional entry *prostituta* defined as *ramera* (whore). There is no definition for the male noun. However, the dictionary contains the interchangeable entry *"prostituto, ta"* (the suffix denotes it can be male or female). This entry refers to the irregular past participle of the verb *prostituir* (prostitute). In the 1956 dictionary, these entries remain with few changes. Both of them keep their definitions, but the entry *prostituta* became an interchangeable word. Most of the changes occurred in the 2014 edition. First, the entry *prostituta* was removed from the dictionary. Second, the entry *"prostituto, ta"* became a definitional entry. And third, the entry *"prostituto, ta"* no longer refers to the past participle, but to the noun, covering both the male and female forms. It also got a neutral gender and a less derogatory definition: a woman or man who engages in sexual acts for money.

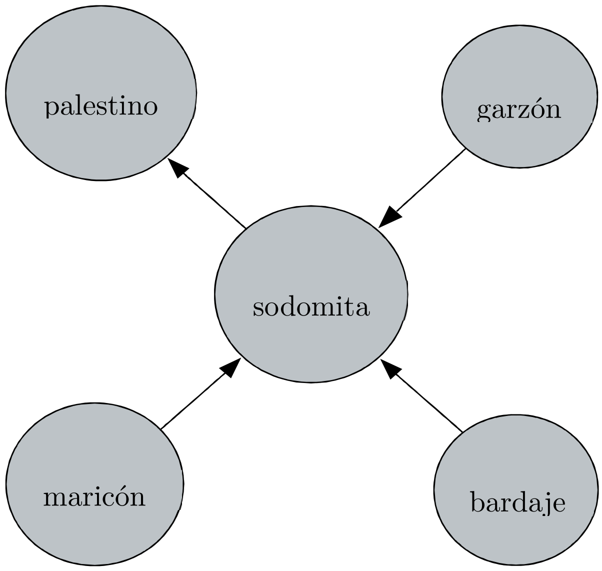
4.3 More complex local changes

The above changes are not particularly surprising (one could guess them, although in the network can be detected automatically!). There are more interesting cases that we think would be difficult or virtually impossible to detect without having a network, and thus, demonstrate in some sense the potentialities of the network methodology. A good example is the evolution in the relationship between the words *sexo* (sex) and *sexual* (sexual) and between *homosexual* (homosexual) and *sodomita* (sodomite).

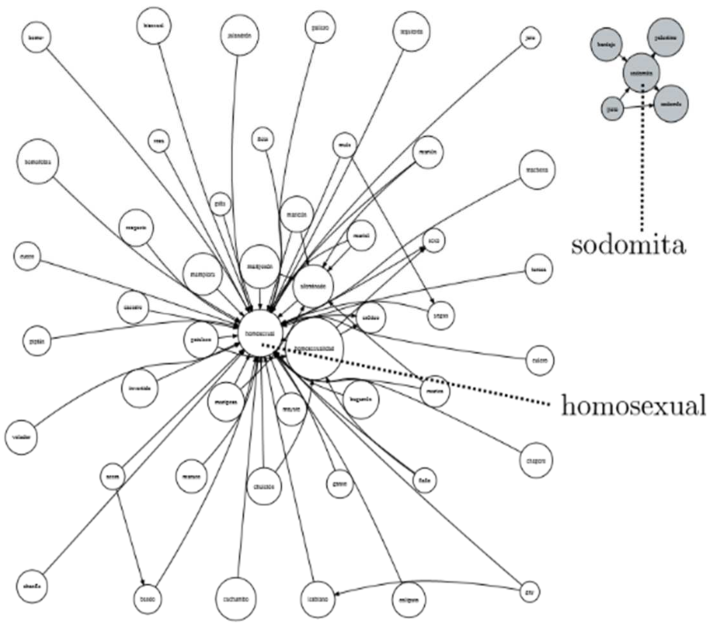
The words *sex* and *sexual* are directly related since the definition of sexual is basically "of or pertaining to sex". However, it is interesting to observe how the relationships between their neighbourhoods change. In 1925 (Figure 7a), the neighbourhood of *sex* is noticeably larger than the neighbourhood of *sexual*; moreover, *sex* was surrounded by biological terms, such as plant, walnut, sweet potato, male, female, hermaphrodite, etc. Later in 1956 (Figure 7b), the size of the neighbourhoods became very similar as *sexual* occurs in more definitions. The neighbourhood of *sexual* expanded to a particular subject. Words such as *sperm*, *egg*, *orgasm*, incorporated *sexual* in their definitions. There are many paths between *sex* and *sexual*, but this edition is the first one to have a word that connects them directly (i.e. there is a path of length 2): *masochism* is defined using both *sex* and *sexual*. Now, in 2014, both neighbourhoods increase their size (Figure 7c), hence their semantic weight. The cloud around *sexual* becomes bigger than that of *sex* and both entries appear where more words connect directly, such as *sexuality*, *venereal*, and *transsexual*.



(a) 1925

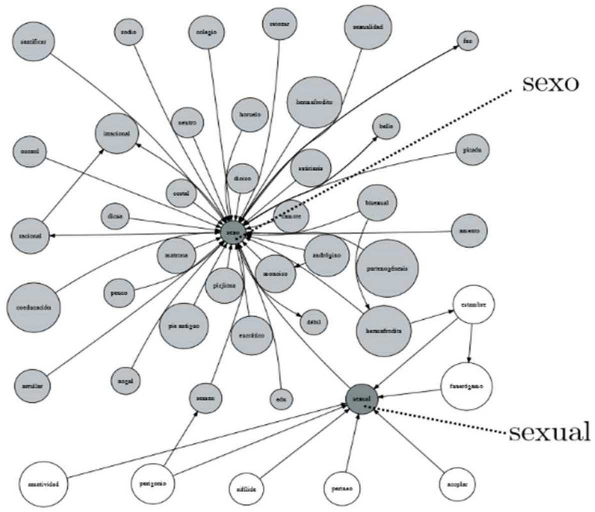


(b) 1956

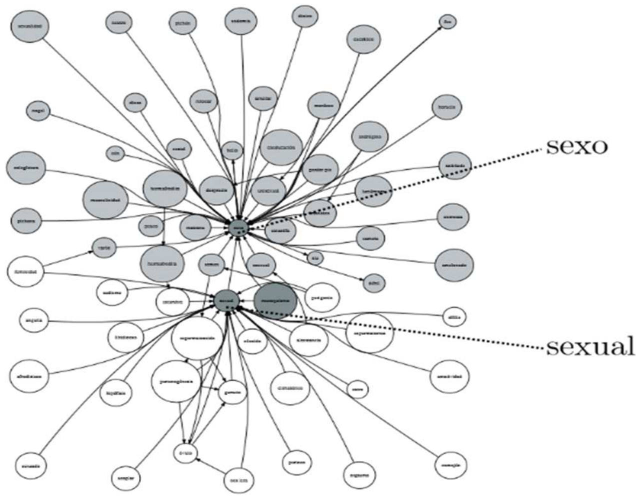


(c) 2014

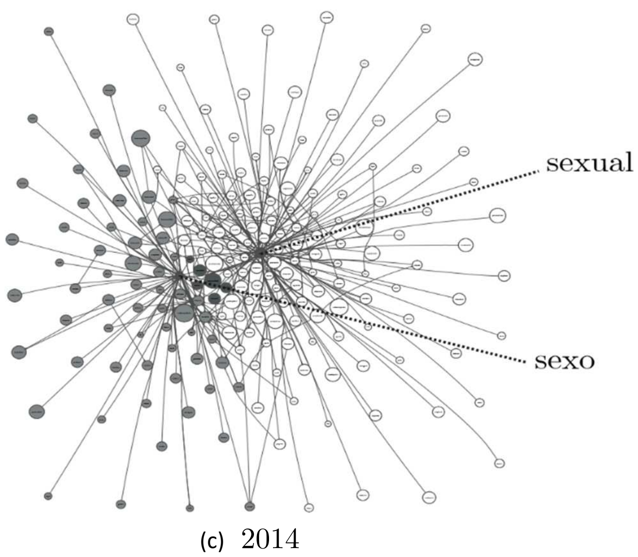
Figure 6: Sub-network around the words *homosexual* (homosexual) and *sodomita* (sodomite).



(a) 1925



(b) 1956



(c) 2014

Figure 7: Sub-network around the words *sexo* (sex) and *sexual* (sexual).

The relationship between *homosexual* (homosexual) and *sodomita* (sodomite) presents a different evolution. In 1925 (Figure 6a), *homosexual* was not defined in the dictionary, while *sodomite* occurred as definitional entry. *Sodomite* covered two concepts: a demonym of an old Palestinian city and a person who engages in sodomy. In 1956 (Figure 6b), the entry *homosexual* was incorporated into the dictionary as a definitional entry. However, it was not a proper definitional entry. It was incorporated as a synonym of *sodomite*, working as a proxy for other words like *homosexuality* to reach *sodomite*. This situation changed in 2014 (Figure 6c), when *homosexual* no longer expressed the meaning of sodomite. It is now defined using concepts such as homosexuality and sexual attraction to persons of the same sex. Its neighbourhood grew considerably; more than 50 words use it in their definitions. Lastly, both entries are not connected anymore. Their concepts diverged. *Sodomite* holds the same meaning since 1925 and *homosexual* evolves from not being in the dictionary, passing to be a synonym of *sodomite*, to become an entry with its own meaning. Last but not least, note that in this analysis the use of neighbourhoods of the network was essential.

5. Related work

Litkowski (1978) was one of the first to state the importance of studying and exploiting dictionary networks, as sources of material for natural language and to unravel the complexities of meaning. He presented three models for representing a dictionary. One based on the relationship *x is used to define y*. The second model incorporates senses of words as nodes. The final model considered the nodes as concepts, having different nodes when words in a definition have more than one meaning.

After Litkowski, there were several investigations about dictionaries and the information that could be extracted from them (Amsler, 1980, 1981; Calzolari, 1984; Chodorow et al., 1985; Calzolari & Picchi, 1988). For example, Picard et al. (2009) aimed to reduce a dictionary to its grounding kernels from which all the other words could be defined. They define a hierarchy of definitional distance and show it correlates with psycholinguistic variables. Levary et al. (2012) studied loops and self-reference in the definition of words. They observed that definitions have a great amount of short loops (length < 6). Muller et al. (2006) presented a method that exploits a directed weighted graph derived from a dictionary to compute distance between two words. The work of Steyvers and Tenenbaum (2005) presented an analysis of the large scale of three types of semantic networks: WordNet (Miller, 1995), word association norms (Nelson et al., 2004), and Roget's Thesaurus (Roget, 1911). They focused on a statistical analysis, concluding that these networks have a small-world structure, characterized by sparse connectivity, short average path lengths between words, and strong local clustering.

Less directly related to our work are lexical databases represented in the form of networks. Built from diverse sources in a manually annotated process, they cover the current use of words and their meanings. WordNet (Miller, 1995) groups words into sets of cognitive synonyms (synsets) and FrameNet (Baker et al., 1998) annotates examples of how words are used in actual texts.

6. Conclusions

This work shows that the study of semantic networks derived from dictionaries could offer insights and tools to study the evolution of the lexicon of a language. We developed in this paper the case of the Dictionary of the Spanish Royal Academy. Among the most relevant findings, is the fact that the network is has a stable structure over the years and is highly resilient. We hypothesize that this is valid for definitional dictionaries in other languages (we tested, although did not present the results here, the case of the English OPTED dictionary). The study presents preliminary evidence that dictionary networks are interesting artefacts and good proxies to study semantic clouds of words and their evolution in a given language.

7. Acknowledgements

The research leading to these results has received funding from under grant CONICYTPCHA/Doctorado Nacional/2015-21150149.

8. References

Amsler, R.A. (1980). The structure of the Merriam-Webster pocket dictionary.

Amsler, R.A. (1981). A taxonomy for English nouns and verbs. In *Proceedings of the 19th annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, pp. 133–138.

Baker, C. F., Fillmore, C. J. & Lowe, J. B. (1998). The Berkeley Framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*. Association for Computational Linguistics, pp. 86–90.

Barabási, A. L. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), pp. 509–512.

Boldi, P. & Vigna, S. (2014). Axioms for centrality. *Internet Mathematics*, 10(3-4), pp. 222–262.

Calzolari, N. (1977). An empirical approach to circularity in dictionary definitions. *Cahiers de Lexicologie Paris*, 31(2), pp. 118–128.

Calzolari, N. (1984). Detecting patterns in a lexical data base. In *Proceedings of the 10th International Conference on Computational Linguistics and 22nd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, pp. 170–173.

Calzolari, N. & Picchi, E. (1988). Acquisition of semantic information from an on-line dictionary. In *Proceedings of the 12th conference on Computational linguistics-Volume 1*. Association for Computational Linguistics, pp. 87–92.

Chodorow, M. S., Byrd, R. J. & Heidorn, G. E. (1985). Extracting semantic hierarchies from a large on-line dictionary. In *Proceedings of the 23rd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, pp. 299–304.

Clark, G. (2003). Recursion through dictionary definition space: Concrete versus abstract words. Technical report, University of Southampton Tech Report).

Dorogovtsev, S. N. & Mendes, J. F. F. (2001). Language as an evolving word web. *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1485), pp. 2603–2606.

Dunne, J. A., Williams, R. J. & Martinez, N. D. (2002). Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences*, 99(20), pp. 12917–12922.

Jeong, H., Mason, S. P., Barabási, A. L. & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, 411(6833), pp. 41–42.

Levary, D., Eckmann, J. P., Moses, E. & Tlusty, T. (2012). Loops and self-reference in the construction of dictionaries. *Physical Review X*, 2(3), p. 031018.

Litkowski, K. C. (1978). Models of the semantic structure of dictionaries. *American Journal of Computational Linguistics*, 81, pp. 25–74.

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11), pp. 39–41.

Muller, P., Hathout, N. & Gaume, B. (2006). Synonym extraction using a semantic distance on a dictionary. In *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*. Association for Computational Linguistics, pp. 65–72.

Nelson, D. L., McEvoy, C. L. & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3), pp. 402–407. URL http://dx.doi.org/10.3758/BF03195588.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2), pp. 167–256.

Newman, M. E., Forrest, S. & Balthrop, J. (2002). Email networks and the spread of computer viruses. *Physical Review E*, 66(3), p. 035101.

Padró, L. & Stanilovsky, E. (2012). FreeLing 3.0: Towards Wider Multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012).* Istanbul, Turkey: ELRA.

Picard, O., Blondin-Massé, A., Harnad, S., Marcotte, O., Chicoisne, G. & Gargouri, Y. (2009). Hierarchies in dictionary definition space. *arXiv preprint arXiv:0911.5703.*

Qi, P., Dozat, T., Zhang, Y. & Manning, C. D. (2018). Universal Dependency Parsing from Scratch. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*. Brussels, Belgium: Association for Computational Linguistics, pp. 160–170. URL https://nlp.stanford.edu/pubs/qi2018universal.pdf.

Roget, P .M. (1911). Roget's Thesaurus of English Words and Phrases. http://www.gutenberg.org/etext/10681. Last accessed 01 July 2017.

Sparck-Jones, K. (1967). Dictionary Circles. Technical report, System Development Corp Santa Monica California.

Steyvers, M. & Tenenbaum, J. B. (2005). The Large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, 29(1), pp. 41–78.