Automatically Identifying Selectional Restrictions of Predicates: A Computational Methodology for Practical Lexicography that Challenges Linguistic Theory

Alexandra Anna Spalek, Marco Del Tredici
Universitetet i Oslo, Universitat Pompeu Fabra
e-mail: a.a.spalek@ilos.uio.no, marco.deltredici@upf.edu

Abstract

In this article we explore a methodology based on unsupervised learning that aims at automatically classifying the arguments of a verb, a classification crucially important in the process of sense isolation in lexicographic projects. More concretely, we strive to exploit the lexical semantic information obtained by combining a clustering algorithm with Word Embeddings and argue that the combination of both represents a fertile methodological ground for automatically classifying the selectional behaviour of a predicate. From the perspective of practical lexicography, our methodology should help to significantly shorten the time invested in isolating context-related senses of a predicate, in our case a verb. In particular, we illustrate, based on a case study of two Spanish verbs, how our methodology can provide a lexicographer with a general picture of a verb's behaviour in a corpus and argue that the clusters we provide automatically encode insightful generalizations that facilitate the process of decision taking of a lexicographer about which senses are most common for a particular predicate, according to the different types of contexts it is most frequently used in.
Keywords: practical lexicography; selectional restrictions of verbs; automatic clustering of arguments; Word Embeddings; *k*-means

1 Introduction

Over the last two decades more and more work has been pointing out that the selectional restrictions of predicates, which semantic theory has assumed to represent compositional patterns of predicate-argument combinations, do not take into account much of the data actually occurring in real language use. Several works, stemming mainly from the discipline of lexicology and lexicography (Bosque 2004, Hanks & Jezek 2008, Hanks 2013, Spalek 2014), have contributed to showing that argument selection paradigms of predicates are much bigger and more diverse than commonly assumed in theoretical approaches to the lexicon. Thus the following Spanish corpus examples of uses of the verb *cortar* "cut", from Spalek (2014), illustrate that it is natural for *cortar* to combine with a complement denoting a physical object, as in (1), along with distinct kinds of events (2), locations (3) and other entities (4):

(1) La policía cortó las cadenas.
the police cut the chains
"The police cut the chains."

(2) Francia corta la entrega de etarras.
France cut the extradition of ETA members
"France stopped the extradition of ETA members."

(3) La policía cortó los puentes que unen sobre el río Vitava la Parte Vieja de Praga con la Parte Pequeña.
the police cut the bridges that unify over the river Vitava the old part of Prague with the small part
"The police blocked the bridges over the river Vitava that unify the old part of Prague with the small part."

(4) Un encargado se refugió en el interior del establecimiento y cortó la energía eléctrica.
a site manager himself hid in the interior of the establishment and cut the energy electric
"A site manager took refuge in the interior of the establishment and cut the electric energy."

At the same time this work has shown that selectional restrictions are not random, but represent concrete clusters, although the actual restriction paradigms look very different from what semantic theory has taught us to think. All this pioneering work accounting for combinatorial paradigms of predicates has been based on manual corpus annotation conducted by linguists and lexicographers. That is, generalizations about relevant argument clusters for a particular predicate have been built on human insights upon the observation of the behaviour of a verb in texts. This theoretic linguistic debate on the nature of selectional restrictions of predicates has its natural counterpart in practical lexicography in the task of classifying the combinatorial behaviour of verbs into patterns that can potentially lead to the classification of relevant senses of a verb (See Hanks 2004 for a methodology). A lexicographer involved in defining the senses of a verb devotes much of his/ her time to classifying prototypical contexts in which the predicate appears, such as the ones illustrated in (1) through (4), in order to identify the senses of this predicate. Parallel to these concerns about selection restrictions coming from theoretical and applied linguistics, evidence from computational linguistics dealing with big data has started to show striking patterns of word uses in natural language, the details of which have only started to be explored more deeply in the last decades.

Under the assumption that patterns of usage of predicates can be analysed and applied in lexicography as a means of discriminating lexical meanings, the present work explores to what extent selectional clusters can be identified automatically. It is worth mentioning here that we are not attempting to identify the meaning of a verb as a word in isolation, but rather to explore a computational method for automatically highlighting potential meanings of the verb associated with its prototypical contexts. Our effort thus focuses on defining a methodology for unsupervised learning of selection restrictions of predicates. From the viewpoint of practical lexicography, we believe that providing a lexicographer with these semantically motivated automatic groupings of verbs' arguments will have major implications for the creation process of lexicographic projects.

From a computational perspective, our work strives to show that human-driven insights on selectional restrictions of predicates can be captured to a significant extent when relying on Natural Language Processing methods like Distributional Semantic Models (DSMs) (Harris 1954). More concretely, we use vector representations of words to explore whether the semantic groupings suggested in manual small-scale works such as Bosque (2004) or Spalek (2014) could be identified automatically on a large scale too. Our work is an advance on previous work in two crucial respects:
1. Based on a case study, we provide a comparison of manually-obtained clusters (Spalek 2014) with clusters gained through unsupervised learning methods.
2. Our case study illustrates that the combination of a clustering algorithm and Word Embeddings serves as a methodology for the automatic identification of selectional restrictions.

2 Methodology for Automatically Detecting Selection Restrictions

While lexicography has been slow in responding to the challenges as well as possibilities presented by big data, computational linguistics has been advancing quickly in analytical data processing methods such as Word Embeddings[1] or clustering algorithms. Clustering and classifications are two common methodologies in computational linguistics. While in classifications the classes to which input instances are assigned are previously defined by humans, and information about these classes is part of the input the algorithm receives, in unsupervised learning clusters are inferred by the algorithm. That means that the main task of unsupervised learning is finding a structure in the data by drawing inferences from datasets consisting of input data without labels and returning clusters of similar objects. (For an overview on clustering algorithms see Aggarwal and Cheng, 2012.) Given that these clustering algorithms have proven to be particularly relevant for the creation of lexical classes (Romeo et al., 2013) as well as for figurative language detection (Del Tredici and Bel, 2015; Shutova and Sun, 2013), we have chosen them as most relevant for our purposes. Our methodology thus takes advantage of the advances made in Word Embeddings and clustering algorithms and tests these methodologies against human-driven practice, as know from practical lexicography.

In the first step of our method, we extracted the list of arguments for our two target verbs, equivalent to the verbs used in the manually annotation of Spalek (2014), namely *cortar*, English 'cut' and *romper*, English 'break'. The present study was limited to the target verbs studied by Spalek (2014) since that study provided concrete results against which our results could be contrasted. Our working corpus, the IULACT corpus (Palatresi 2009), was limited only to the general corpus compiled from Spanish newspapers. This was done to keep the composition of the corpus as similar as possible to the corpus used in Spalek (2014).

Secondly, for the obtaining of automatic groupings of arguments with high semantic similarity, analogue to the manual groupings in other studies (Bosque, 2005 or Spalek, 2014), we used vector representations proceeding from Distributional Semantic Models[2]. In general, DSMs employ vectors

[1] Word Embeddings are vector representations of words: each word is associated with a vector, which corresponds to a point in a multidimensional space. The distance between the points in space is a function of the similarity of the words. Mathematically that means that the closer two vectors, the more similar are the words they represent.

[2] DSMs are based on the distributional hypothesis (Harrys 1954), whereby semantically similar words tend to have similar contextual distributions

that keep track of the contexts in which target terms, in our case *cortar* and *romper,* represent the meaning of words as points in a vector space. Employing a DSM thus allowed us to measure the similarity in the meaning of different words using geometric techniques such as cosine similarity (Baroni, Dinu and Kruszewski 2014). As an example, the word *cristal,* ("glass"), turned out to have higher cosine similarity with the word *ventana*, ("window"), than with the word *silencio*, ("silence"). Among several available DSMs, we chose Word Embeddings (WEs), a neural-network based vector representation in which words are embedded into small-size vectors (Mikolov 2013). This choice was based on the observation that WEs have been attested to be extremely efficient in mapping meaningful syntactic and semantic information of words. In addition, recent works (Baroni, Dinu and Kruszewski 2014; Del Tredici and Bel 2014) have argued that WEs, in contrast to other DSMs, are currently considered to be most appropriate for word representations. For our study, WEs were implemented on the Spanish section of the freely available Wikicorpus (Reese 2010), which contains around 120 million words and is annotated with lemma, part of speech information and word senses. Adopting common practice, we ignored words that were rare (less than 150 occurrences) in the training corpus (Mikolov 2013). Our clustering algorithm was run on this corpus in order to create groups of similar arguments for the verbs *cortar* and *romper*. We considered two different options to perform this task: *k*-means and hierarchical clustering.

K-means is a partitioning method widely used in Natural Language Processing (NLP) for its effectiveness and its low computational complexity. The algorithm takes as input a number of objects and returns as output a set of clusters, where similar objects are assigned to the same cluster, and each object can belong to just one of these clusters (hard clustering). *K*-means is said to be *partitional* (or *flat*) because, unlike hierarchical clustering (see below), it returns a flat set of clusters in which clusters are not organized in a hierarchical structure. At the core of the algorithm is the idea of *centroid*, which can be defined as the central point of a cluster. Before running the algorithm the user has to define the number *k* of centroids, which corresponds to the number of clusters returned as the output. With the *k* value as input, the algorithm randomly instantiates *k* centroids in the semantic space (see the grey crosses in Figure 1). The final objective of the algorithm is to adjust the position of the centroids (dotted grey lines in Figure 1) so that they end up in a position where all the points in each cluster are as near as possible to their centroid (black crosses in Figure 1). This is achieved by performing many iterations, and adjusting the position of the centroid during each iteration. The algorithm stops when no more adjustments are possible. When working with *k*-means, the number of clusters for the output has to be defined before running the algorithm. Given that in Spalek (2014) the number of clusters had already been defined, in order to perform a fair evaluation, we set the value of *k* for each target verb equal to the number of clusters defined there: $k = 11$ for *romper* and $k = 14$ for *cortar*. However, this is not always possible. In fact, this is a well-known problem in the NLP community to which many diverse solutions have been proposed. The most common one is based on the usage of external metrics, which are employed to assess the quality of the clusters in different clustering solutions. The basic idea is to perform *k*-means with several values of *k*, and then choose the value for which the best set of clusters has been obtained. Despite the fact that this kind of methodology requires a high number of iterations, it is still feasible given the low computational complexity of the algorithm. As an external measure to evaluate the quality of the clustering, we propose the Silhouette score (Rousseeuw, 1987), a widely employed metric for the interpretation and validation of clustering results which gives in output a value between –1 and 1, where the maximum value (1) indicates that objects in the same clusters are highly similar (high *cohesion*) and at the same time, highly different from those in other clusters (high *separation*).

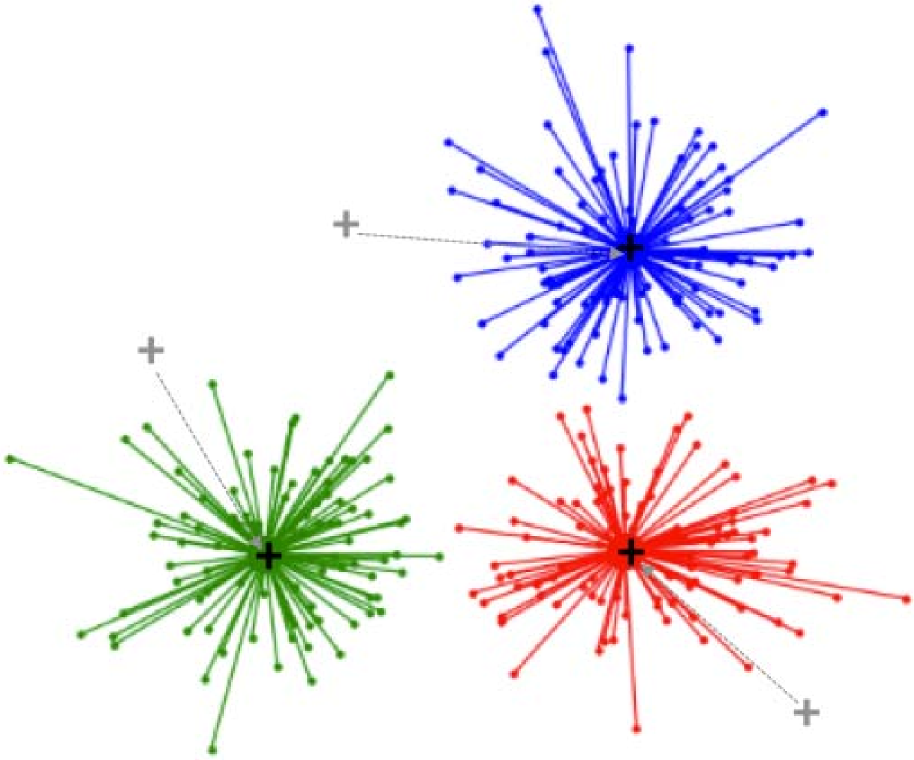
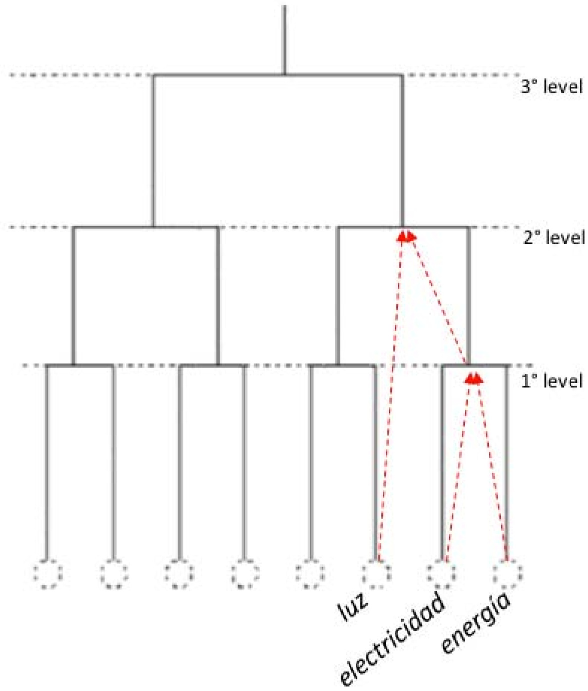


Figure 1 An example of *k*-means clustering, where *k*=3. Grey crosses indicate the position of the centroids before the first iterations. After many iterations, centroids are moved (dotted lines) untill they end up in a position where all the points in each cluster are as near as possible to their centroid.

Hierarchical clusters are an alternative option, differing from the flat output returned by the *k*-means algorithm. These give as output a hierarchical structure (also called "dendogram") that provides a wider overview on how the objects in the dataset are related to each other, showing the relations of hyponymy/ hyperonym among them. This kind of algorithm implements a bottom-up approach, in which each word is initially considered as a single cluster. The algorithm performs many iterations and during each iteration, the two most similar clusters are merged in a new cluster, which thus creates a hierarchy. Iterations are repeated until only one cluster, including all the words, is created at the top of the hierarchy. The higher the level in the hierarchy, the smaller the number of clusters and, consequently, the larger the size of the clusters. As a result of applying this algorithm, we find that, depending on their position in the hierarchy, clusters differ in their degree of granularity. Clusters at the bottom of the hierarchy are fine-grained and thus include only highly similar vectors (e.g. *electricidad* 'electricity' and *energía*, 'energy' with respect to the verb *cortar*, English *cut*). Towards the top, clusters are more coarse-grained. This means that the vectors they include are still related but less similar (e.g. *electricidad* 'electricity' and *luz* 'light', see Figure 2). Compared to *k*-means, hierarchical clustering has the advantage that the number of clusters in the output does not have to be defined before running the algorithm. Thus, it is common practice to observe the resulting hierarchy and choose the best level of granularity needed for the specific purposes of the research, and select the set of clusters at that level. However, hierarchical clustering presents an important disadvantage compared to *k*-means, namely its computational complexity: this simply means that this algorithm requires a great deal of computing power, which is not always available, and which, even if available, makes the algorithms slow for large datasets. As for the present research, it should be noted that despite the differences between the two algorithms, if both are employed with the appropriate setting on the same dataset, they are expected to output similar results. That is, what we expect is that the best flat set of clusters returned by the *k*-means algorithm will approximately correspond to the most suitable level of the hierarchy for our specific research purpose. Thus, considering the two methodologies as equivalent, our decision on which one of the two to pick favoured computational



simplicity and we considered *k*-means a better choice.

Figure 1 Simplified example of a hierarchy obtained with hierarchical clustering. Words to be clustered are the *leaves* of the tree (in this case, the labelled leaves are *luz* 'light', *electricidad* 'electricity' and *energía*, 'energy'). During the first iteration the algorithm creates small clusters including the couple of words with the most similar vector representation (e.g. *electricidad* and *energía*). The output of the first iteration is the 1st level of the hierarchy. Cluster at this level are very fine-grained. At the second iteration of the algorithm, a cluster in the 1st level is merged with the most similar cluster at the same level in order to create a larger cluster (in the example, the cluster including *electricidad* and *energía* is merged with the one including *luz*). The output of the iteration is the 2nd level of the hierarchy. Clusters at this level are more coarse grained, meaning that the similarity among the words they include is lower compared to the similarity of the words in clusters at the 1st level. The same procedure is repeated for subsequent levels of the hierarchy, until all the words are included in the same cluster (top of the hierarchy).

3 Results and Discussion

In the previous section we presented two different algorithms to perform clustering (*k*-means and hierarchical clustering) and argued for employing *k*-means, setting the value of *k* for each target verb equal to the number of clusters defined in Spalek (2014). In this section we present the results of applying our methodology and critically evaluate them, contrasting them with the results stemming from a manual clustering process, such as in Spalek (2014).

Firstly, our study confirms the observations of previous studies (Bosque 2004, Hanks and Jezek 2008, Spalek 2014), namely that abstract entities denoting arguments effectively outnumber physical objects as direct objects of change of state verbs, and that this is the case even when extending the sample to a large corpus, as we did for this study. More specifically, for *romper* we found that 16.2% of typical arguments were physical objects, while 83.8% corresponded to abstract objects. For *cortar*, physical objects represented 30.7% and abstract objects 69.3%. Our results thus contribute to emphasise the highly different percentages of abstract and concrete arguments as typical combinatorial patterns of verbs. The fact that reaching out to big data clearly shows that verbs so significantly and naturally select for all kind of abstract-object-denoting nouns is not entirely new but is challenging, in as far as the assumptions usually made in theoretical linguistics that verbs refer to events in the physical world has led many theoretical linguistic models to ignore much of the real data. At the same time this fact provides the lexicographer with a very general insight that transitive verbs have crucially two meaning options: 1) they denote physical events and 2) they denote abstract events.

Secondly, taking under examination some well-studied data from previous work (Spalek, 2014), our study shows that it is possible to obtain semantically consistent clusters of verb arguments by applying our methodology. More concretely, the methodology presented here allows us to automatically identify semantic clusters of arguments that potentially pick out distinct meanings of a predicate. For the verb *cortar* "cut" the automatically found clusters of arguments correspond to 11 out of the 15 argument groups in Spalek (2014): [Physical Object], [Food], [Body Part], [Location], [Continuous Stream], [Event], [Time Interval], [Process], [Relation], [Human | Group], [Feeling]. The missing three clusters were either empty or irrelevant. For *romper,* English "break", the automatically generated groupings of arguments coincided with 10 out of the 11 identified in Spalek (2014): [Physical Object], [State], [Norm], [Doctrine], [Relation], [Act], [Process], [Time Interval], [Group], [Human | Group]. Comparing our clusters in more detail to the clusters found in previous work, our semantic groupings of arguments are in fact similar in content to what has been identified in the manual annotation of Spalek (2014), although some significant differences have to be mentioned: when running statistical purity tests, comparing the lexical clusters found in Spalek (2014) to the ones automatically identified by the clustering algorithm, we obtain a purity of 0.6. That means that the manual classification and the automatic clusterings are indeed significantly different. A relevant observation at that point, however, is that the groupings in Spalek (2014) – our comparison class – are not hard categories, but could actually be redefined in a different way. A good example to explain the differences in the classification is the noun *silencio*, ("silence"). In the classes of Spalek (2014) *silencio* was classified as a member of the semantic group [State]. Our automatic cluster, however, classified *silencio* together with other words denoting processes ([Process]). This difference in classification, rather than representing an incorrect clustering, seems to point out a problem proper to the category of abstract nouns. These nouns seem to group into clusters in a relatively lax way (See also Hands and Jezek (2008) for a similar observation) and can therefore naturally be classified in different clusters. Nouns referring to physical objects, in contrast, seem to cluster together very consistently, be it in the manual annotation or in the automatic classification. This observation is supported by the fact that the percentage of purity of the cluster identified by the clustering algorithm that groups together physical objects includes no errors whatsoever, while other clusters did include some misclassified elements. Furthermore, for many other arguments denoting abstract events and entities we also attribute the relatively low purity to the fact that Spalek (2014) used a different corpus, and many of the integrants of the clusters we have identified automatically are not semantically irrelevant to the clusters they are members of, but rather just absent from the sample manually annotated by Spalek (2014). When reviewing the semantic quality of our automatically identified clusters, we still find that they are consistent in grouping semantically similar words. For this reasons we take the clusters created by the algorithm as an alternative to Spalek (2014), rather than considering them to be wrong.

Most importantly, however, the argument clusters identified automatically in our work can effectively be correlated with the examples analysed by Spalek (2014) ((1) through (4)). Thus example (1) clearly corresponds to our cluster of [Physical Objects], example (2) represents a member of our semantic group of [Events], example (3) represents the group we have labelled [Location], and example (4) represents our cluster of [Continuous Stream]. This is to say the automatically identified clusters are significant, since they clearly help to identify different senses of the verb, which is of interest for lexicographic praxis. More concretely, a look at the definition of *cortar* in the dictionary of the Spanish Royal Academy (DRAE) reveals that the clusters identified can actually be correlated with different senses provided in the DRAE. Thus, the clusters [Physical Object] and [Body Part] can be correlated with the separation sense of *cortar* (sense 1), the [Food] cluster can be correlated with the division-into-pieces sense (sense 5), the [Location] cluster can be associated with the blocking sense (sense 9), the [Event] and [Process] cluster can be related to the meaning of interruption (sense 16), the [Time Interval] cluster points at the sense of shortening (sense 8) and the [Human Group] can be correlated with the meaning *cut off a speaking person* (sense 10). Finally our algorithm also identified another minor cluster that grouped together nouns referring to feelings that points to the sense of becoming shy (sense 27).

For a critical evaluation of our methodology and how relevant are the automatically identified clusters we can say that, in the current state of the methodology, automatic clustering is both outperforming and underperforming to a certain extent. It outperforms, with respect to the current definition of *cortar* in DRAE, in that it found some clusters that are not present in the DRAE. Thus our cluster [Relation] points to a very common use of *cortar* referring to the break-up of relations, as a sense that other Spanish dictionaries do include. Similarly the [Continuous Stream] cluster consists of members such as *electricidad* “electricity”, *energía* “energy” or *gas “*gas*”* for which *cortar* refers to events of cutting off the supply of any of them. This is also a sense commonly used in other dictionaries while lacking from the DRAE. At the same time our methodology does not seem to be exhaustive. For instance, no cluster corresponding to *substances* was found. Although not being one of the most important senses, many Spanish dictionaries, as well as Spalek (2014), include this sense as an event of *cortar* denoting the dilution of a substance. In the DRAE this use corresponds to senses 14 and/ or 15.

Summing up the results so far, it seems that, despite several shortcomings, our methodology for the automatic identification of selection restriction of verbs seems to provide the lexicographer with a first general overview of the behaviour of a predicate, an essential basic step for a lexicographic project that starts from scratch. Our prediction here is that implementation of our methodology would improve and speed up the process of lexicographic work based on corpora, for it automatically picks out the typical combinatorial paradigms of a predicate. This valuable first general picture can then be further refined by information from introspection or comparison to other lexicographic projects to fully account for all possible senses of a predicate.

4 Some Relevant Remarks for the Theory of Word Meaning

The results emerging from our study also serve as insightful input for theoretical approaches to the lexicon. Given that much of the linguistic literature on verb meaning in the tradition of theoretical linguistics has focused on analysing verb uses when they describe events in the physical world, theoretical linguists has ended up postulating selectional restrictions limited to events describing the physical world only. Consequently, the verb's combinatorial capacity has been restricted by default to nouns denoting physical entities. And though referential approaches to semantics have proven very successful at providing meaningful analyses for a wide range of natural language data, lexical meaning has largely eluded insightful treatment when verbs refer to events that go beyond the physical world. As a matter of fact, however, our large-scale clustering of arguments of a verb illustrates that the range of combinatorial contexts of a verb vastly surpasses the domain of physical entities. This fact has previously become evident at a smaller scale with the publication of the combinatorial dictionary of Spanish, REDES (Bosque 2004), which provides the user with the lexical classes ("argument clusters" in our terms) of semantic notional groupings of the inventory of arguments attested for a particular verb. For the Spanish *cortar*, REDES lists 13 lexical classes of arguments. This is very close to the number of groupings identified in Spalek (2014) as well as the number of clusters that turned out to be relevant in our case study.

From a theoretical point of view our work contributes to evidence that just about every simple transitive verb expresses a wide range of predicates depending on the variety of direct objects it can take. Probably not surprisingly for lexicographic praxis, nor from the computational perspective, but new to theoretical approaches to lexical meaning, only two lexical classes in REDES entry for *cortar* correspond to physical objects, while the rest of the classes are distinct kinds of abstract arguments. Thus our enterprise of finding a methodology that lets the company of a predicate tell us what senses a predicate can have reinforces the conclusion of theoretical studies that there is an urgent need to take seriously big data proceeding from corpora when elaborating any models of word meaning.

5 Conclusions

Our study has united insights from theoretical linguistics with the praxis of lexicography as well as current methods in computational linguistics to explore an innovative methodology that enhances lexicographic work by automatically extracting the compositional patterns of a predicate in real text. Using a cluster algorithm and Word Embeddings has turned out to be a fertile method for providing quick insight into the possible variety of senses of one predicate in direct relation to its arguments. We thus hope to have suggested a practical way to use unsupervised learning methods to deal with big data in lexicographic projects with the potential for significant saving of time.

6 References

Aggarwal, Charu C., and Cheng Xiang Zhai (2012)."A survey of text clustering algorithms." Mining Text Data. Springer US, 77-128

Bosque, I. (2004). REDES *Diccionario combinatorio del Español Contemporáneo.* Madrid: Ediciones SM.

Baroni, Marco, Georgiana Dinu, and Germán Kruszewski (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. I*n Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics.* 238–247. Baltimore, Maryland, USA, June 23-25 2014.

Del Tredici, Marco, and Núria Bel. (2015). A Word-Embedding-based Sense Index for Regular Polysemy Representation. In P*roceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing.* Denver, USA

Hanks, P. (2004). *Corpus Pattern Analysis. Proceedings of the XIII EURALEX International Congress*, ed. Geoffrey Williams and Sandra Vessier, 87-97. Lorient, France.

Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations*. Cambridge, MA: MIT Press.

Hanks, P. & Jezek, E. (2008). Shimmering lexical sets. In *Proceedings of the XIII EURALEX International Conference*, ed. Elisenda Bernal and Janet DeCesaris, 391–402. Barcelona, Spain.

Harris, Zellig S. (1954). *Distributional* structure. Word 10.2-3 (1954): 146-162.

Levy, Omer & Yoav Goldberg (2014). Dependency based word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Vol. 2, Baltimore, USA.

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

Mikolov, Tomas, et al. (2013). *Efficient estimation of word representations in vector space.* arXiv preprint arXiv, 1301-3781.

Palatresi, Jorge Vivaldi (2009). Corpus and exploitation tool: IULACT and bwanaNet. In: A survey of corpus-based research [Electronic resource], p. 224-239.

Real Academia Española (2014). "cortar" *Diccionario de la lengua española* (23.ª edición). Madrid: Espasa. Consulted 13th de april 2016.

Reese, Samuel, et al. (2010). Word-sense disambiguated multilingual Wikipedia corpus. *7th International Conference on Language Resources and Evaluation*. Malta.

Romeo, Lauren, Sara Mendes, and Núria Bel. (2013). "Using qualia information to identify lexical semantic classes in an unsupervised clustering task." arXiv preprint arXiv:1303-2449 .

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Shutova, E. and Lin Sun (2013). "Unsupervised Metaphor Identification Using Hierarchical Graph Factorization Clustering." HLT-NAACL.

Spalek, A. (2014). Verb Meaning and Combinatory Semantics: A Corpus-Based Study of Spanish Change of State Verbs. Barcelona: Universitat Pompeu Fabra.

Acknowledgements

This research has been supported by the SynSem project at the Faculty of Humanities at Universitetet i Oslo.