Distributional Semantics Today Introduction to the special issue

Cécile Fabre, Alessandro Lenci

Distributional Semantics Today
Introduction to the special issue

Cécile Fabre* — **Alessandro Lenci****

* *University of Toulouse, CLLE-ERSS*
** *University of Pisa, Computational Linguistics Laboratory*

ABSTRACT. *This introduction to the special issue of the* TAL *journal on distributional semantics provides an overview of the current topics of this field and gives a brief summary of the contributions.*

RÉSUMÉ. *Cette introduction au numéro spécial de la revue* TAL *consacré à la sémantique distributionnelle propose un panorama des thèmes de recherche actuels dans ce champ et fournit un résumé succinct des contributions acceptées.*

KEYWORDS: *Distributional Semantic Models, vector-space models, corpus-based semantics, semantic proximity, semantic relations, evaluation of distributional ressources.*

MOTS-CLÉS : *Sémantique distributionnelle, modèles vectoriels, proximité sémantique, relations sémantiques, évaluation de ressources distributionnelles.*

1. Introduction

Distributional Semantic Models (DSMs) have been the focus of considerable research over the past twenty years. The use of distributional information extracted from corpora to compute semantic similarity between words has become a very common method in NLP. Its popularity is easily explained. In distributional models, the meaning of words is estimated from the statistical analysis of their contexts, in a bottom-up fashion: requiring no sources of knowledge other than corpus-derived information about word distribution in contexts, it provides access to semantic content on the basis of an elementary principle which states that semantic proximity can be inferred from proximity of distribution. It gives access to meaning in usage, as it emerges from word occurrences in texts. Distributional semantics based on vector space models has benefited from the availability of massive amounts of textual data and increased computational power, allowing for the application of these methods on a large scale. Today, the field has reached maturity: many experiments have been carried out on different languages, several survey articles have helped to consolidate the concepts and procedures used for distributional computations (Turney and Pantel, 2010; Baroni and Lenci, 2010), various distributional models and evaluation data are now available. Still, many issues remain open to gain a better understanding of the type of information that is induced by these methods and to extend their use to new applications and new linguistic phenomena.

In recent years, much research effort has focused on optimization methods to handle massive corpora and on the adjustment of the many parameters that are likely to have impact on the quality and nature of semantic relations. A second important issue relates to the relevance of distributional semantic information for a large number of tasks and applications. Finally, in the last few years, research has also focused on a better understanding of the semantic information that is conveyed by these models. Before presenting the papers that appear in this special issue of the *TAL* journal dedicated to distributional semantics, this introduction provides an overview on these different topics.

2. Principles and Methodology of the Construction of DSMs

2.1. *The Distributional Hypothesis*

Distributional semantics is grounded on the Distributional Hypothesis: *similarity of meaning correlates with similarity of distribution*. Zellig Harris is usually referred to as the theoretical and methodological source of research for distributional semantic models (Harris, 1954). In fact, he considered distributional method the only viable scientific approach to the study of linguistic meaning. In his later works, he designed a method to classify words on the basis of the contexts they share in a given corpus, through the careful collection and analysis of dependency relations involving operators and arguments (Harris, 1991). What was clearly asserted in Harris' original method was the fact that such inductive semantic classifications reflected the

use of words in specific corpora. The approach was set in the context of the theory of sublanguages, based on the assumption that only corpora from restricted domains could guarantee the possibility to build up clear-cut semantic categories (Habert and Zweigenbaum, 2002).

Since the 1960s, several implementation of the Distributional Hypothesis have been carried out for the automatic constructions of thesauri for machine translation and information retrieval (Grefenstette, 1994). A crucial contribution to distributional semantics has indeed come from the vector space model in information retrieval (Salton *et al.*, 1975), resulting in successive improvements to the original methodology with respect to the nature of data and the mathematical formalization, thereby boosting its spread in computational linguistics. In the last twenty years, the possibility to apply the method on a much larger scale to huge corpora has imposed the distributional approach as the default approach to semantics in NLP.

2.2. *Design of Distributional Semantic Models*

In DSMs, words are represented as vectors built from their distribution in contexts, and similarity between words is approximated in terms of geometric distance between their vectors. The standard organization of DS systems is usually described as a four-step method (Turney and Pantel, 2010): for each target word, contexts are collected and counted and a co-occurrence method is generated; raw frequencies are then usually transformed into significance scores that are more suitable to reflect the importance of the contexts; the resulting matrix tends to be very large and sparse, requiring techniques to limit the number of dimensions. Finally, a similarity score is computed between the vector rows, using various similarity measures. DSMs have many design options, due to the variety of parameters that can be set up at each step of the process and may affect the results and performances of the system.

2.2.1. *Parameters*

A corpus-based semantic model reflects the semantic behaviour of words in use. It is thus by definition highly dependent on the type of corpus that is being analyzed. There has been a clear shift from the treatment of middle-sized specialized corpora for the acquisition of distributional thesauri in the 90's (Grefenstette, 1994; Nazarenko *et al.*, 2001), to the compilation of corpora as large as possible, often heterogeneous in genre and domain. Newspaper and encyclopedic articles (Peirsman and Geeraerts, 2009), balanced reference corpora such as the BNC (Sadrzadeh and Grefenstette, 2011), very large corpora obtained from the web (Agirre *et al.*, 2009), or any combination of the former (Baroni and Lenci, 2010) have been used. The trend to use huge corpora is mainly motivated by the joint need of increasing the coverage of distributional lexical resources while reducing data-sparseness, which is known to negatively affect the performance of DSMs.

The definition of contexts is another crucial parameter in the implementation of the systems. Three types of linguistic environments have been considered (Peirsman and Geeraerts, 2009): in document-based models, as in *Latent Semantic Analysis* (LSA) (Landauer and Dumais, 1997), words are similar if they appear in the same documents or in the same paragraphs; word-based models consider a "bag-of-word" window of collocates around the target words (Lund and Burgess, 1997; Sahlgren, 2008; Ferret, 2013); syntax-based models are closer to Harris' approach as they compare words on the basis of their dependency relations (Curran, 2004; Padó and Lapata, 2007; Baroni and Lenci, 2010). Word-based models have an additional parameter represented by the window size (from a few words to an entire paragraph), while syntax-based models need to specify the type of dependency relations that are selected as contexts (Baroni and Lenci, 2010; Peirsman *et al.*, 2007). Some experiments suggest that syntax-based models tend to identify distributional neighbors that are taxonomically related, mainly co-hyponyms, whereas word-based models are more oriented towards identifying associative relations (Van de Cruys, 2008; Peirsman *et al.*, 2007; Levy and Goldberg, 2014). However, the question whether syntactic contexts provide a real advantage over "bag-of words" models is still open. On the other hand, a more dramatic difference exists with respect to document-based models, which are strongly oriented towards neighbors belonging to loosely defined semantic topics or domains (Sahlgren, 2006).

Other parameters have received particular attention: weighting scores and similarity measures. A wide range of setting exists for both parameters (Curran, 2004; Bullinaria and Levy, 2007), but nowadays the most common practice is to use Positive Pointwise Mutual Information as weighting scheme and cosine as similarity measure, which are typically credited for granting the best performances across a wide range of tasks (Turney and Pantel, 2010).

Vectors in the co-occurrence matrix provide an *explicit* representation (Levy and Goldberg, 2014) of the lexeme distribution in contexts. Each vector dimension in fact represents a specific context in which the target word has been observed. Explicit co-occurrence vectors are huge and sparse. Techniques are therefore used to reduce their dimension and limit computational complexity. The most common approach consists in mapping the original sparse matrix into a low-dimensional dense matrix with methods such as Singular Value Decomposition (Landauer and Dumais, 1997), Non-Negative Matrix Factorization (Van de Cruys, 2010), and Latent Dirichlet Allocation (Blei *et al.*, 2003). Crucially, the dimensions of the reduced vectors no longer correspond to explicit contexts, but rather to "latent" semantic dimensions implicit in the original distributional data. Matrix reduction techniques smooth unseen data, remove noise and exploit redundancies and correlations between the linguistic contexts, thereby improving the quality of the resulting semantic space (Turney and Pantel, 2010). A popular as well as effective alternative to matrix reduction is Random Indexing (Sahlgren, 2006): instead of reducing a previously constructed matrix, low-dimensional representations are incrementally built by assigning each word a random vector that is summed to the vectors of the co-occurring words.

Much research has been dedicated to the investigation of the impact of some or all these parameters on the performance of DSMs systems in a variety of tasks. The most recent and comprehensive studies are those of Lapesa and Evert (2014) and Kiela and Clark (2014). They investigate a very large set of parameters, including type of corpus, use of stemming and lemmatization, type of contexts (dependency vs co-occurrence, direction and size of the window), weighting scores, similarity measures, dimensionality reduction techniques. These experiments provide a very useful presentation of the best configurations according to the type of semantic task.

2.2.2. *Count vs. Prediction Models*

The DSMs we have just described use a *count-based* approach to build distributional representations: corpus co-occurrences are first counted, then weighted and finally optionally reduced to dense vectors. Recently, a new family of *prediction-based* DSMs has appeared: neural network algorithms directly create dense, low-dimensional word representations by learning to optimally predict the contexts of a target word (Mikolov *et al.*, 2013a). These representations are also typically referred to as *embeddings*, because words are embedded into a low-dimensional linear space of latent features. Various types of "linguistic regularities" have been claimed to be identifiable by embeddings (Mikolov *et al.*, 2013b). For instance, the fact that *king* and *queen* have the same gender relation as *man* and *woman* is represented in their embeddings offsets, so that the vector of one word (e.g. $queen$) can be recovered by the representations of the other words by simple vector arithmetics (i.e., $king - man + woman$). Moreover, prediction-based models have been shown to outperform count-based ones in various semantic tests (Baroni *et al.*, 2014)

Despite their increasing popularity, the question whether embeddings are really a breakthrough with respect to more traditional methods is far from being set. For instance, the same linguistic regularities captured by embeddings are also captured by explicit count-based models (Levy and Goldberg, 2014). When parameters of the latter are carefully tuned, no significant difference is observed in the performance between count and prediction-based models (Levy *et al.*, 2015). It is possible that future research will able to show some clear advantage for embeddings, but for the time being the two approaches do not substantially differ for the type of semantic aspects they are able to address. They are just alternative ways to build distributional representations.

3. Evaluation of DSMs

The classical dichotomy between *intrinsic* and *extrinsic* modes of evaluation in NLP applies to DSMs as well. Intrinsic evaluations aim at measuring the quality of the resource in itself, by confronting it with human evaluation or with similar semantic resources that can be used as gold standards. Extrinsic evaluations measure the specific contribution of the resource to enhance the performance of a system in which it is integrated.

The intrinsic evaluation of the DSMs has been conducted through the comparison to various lexical resources, such as the TOEFL synonym detection task (Landauer and Dumais, 1997), specialized thesauri (Grefenstette, 1994), wordnets (Curran and Moens, 2002; Padró *et al.*, 2014; Anguiano and Denis, 2011), synonym dictionaries (Van der Plas *et al.*, 2011). Intrinsic evaluation of DSMs is a complex issue for various reasons. First of all, DSMs capture a very broad notion of semantic proximity (cf. section below). Therefore, there is an inevitable mismatch between DSMs results and resources that focus on specific, classical lexical relations, such as synonymy dictionaries, thesauri and wordnets. A second kind of potential mismatch is due to the fact that DSMs results reflect the specificities of the corpus and as such they can identify potentially relevant semantic relations and yet missing in general-purpose resources. It is indeed difficult, perhaps impossible, to assess the validity of a semantic relation out of context (Muller *et al.*, 2014). In order to overcome such limitations, a number of resources specifically geared towards DSM evaluation have been developed, mostly for English. One of the most popular gold standard is WordSim-353 (Finkelstein *et al.*, 2002), with 353 word pairs rated by human judgments. A multilingual version of this dataset has also been recently released (Leviant and Reichart, 2015).

Regarding extrinsic evaluation, the use of distributional features is useful each time there is a need to compute similarities between words or longer stretches of text. Several experiments have been dedicated to the use of distributional resources in information retrieval to compute query similarity (Alfonseca *et al.*, 2009; Claveau and Kijak, 2015). In the lexical substitution task (McCarthy and Navigli, 2007), a DSM is used to compute potential substitutes before the disambiguation process (Fabre *et al.*, 2014). Distributional similarity is also used as a cue to determine the predominant sense of a word in a corpus (McCarthy *et al.*, 2007). DSMs have proved efficient in even more complex NLP applications such as textual entailment or summarization (Cheung and Penn, 2013). Word embeddings have also been successfully used to improve Semantic Role Labeling and Named Entity Recognition (Collobert and Weston, 2008).

4. The challenges for DSMs

Critics have been regularly addressed to DSMs, even by researchers involved in the field: the bottom-up approach to meaning pursued by distributional semantics is very practical in terms of processing, but it is an open issue whether statistical co-occurrences alone are enough to address deeper semantic questions or just provide a shallow proxy of lexical meaning (Sahlgren, 2008; Lenci, 2008; Koller, 2015).

The Distributional Hypothesis is a claim about semantic similarity, which DSMs measure with proximity in vector spaces. However, semantic similarity is itself a very vague notion, ranging from similarity between words to similarity between relations (Turney, 2006; Baroni and Lenci, 2010; Turney, 2013). It is also necessary to distinguish semantic similarity *stricto sensu* (also called *attributional similarity*), as a relation between words sharing similar semantic features, such as as *car* and *van*, from

the *semantic relatedness* of words that are strongly associated, like *car* and *wheel* (Budanitsky and Hirst, 2006; Agirre *et al.*, 2009). These two types of similarities have very different semantic properties. Yet, they are hardly distinguished by DSMs. Even gold standards like WordSim-353 are populated with semantically related pairs (Agirre *et al.*, 2009). In order to address this issue, the dataset SimLex-999 has been recently developed in order to specifically evaluate DSMs' ability to capture semantic similarity rather than semantic relatedness (Hill *et al.*, forthcoming).

An additional problem is that both semantic similarity and semantic relatedness are cover terms for very different types of lexical relations. For instance, both synonyms, co-hyponyms and even antonyms can be said to be semantically similar because they share a high number of features. Semantic relatedness includes meronymy, locative relations, up to topical and other non-classical relations (Morris and Hirst, 2004). This large and graded notion of relatedness is both useful and problematic for NLP applications, because it is very difficult to draw a clear limit between relevant and non-relevant associates (Sahlgren, 2008; Ferret, 2013). In general, the distributional neighbors identified by DSMs have very different semantic relations with the target, suggesting that DSM provide quite a coarse-grained representation of lexical meaning. The BLESS (Baroni and Lenci, 2011) and the most recent EVALution (Santus *et al.*, 2015) datasets have specifically been designed to test the ability of DSMs to discriminate different types of relations, which represents an important area of research in distributional semantics (Van der Plas and Tiedemann, 2006; Lenci and Benotto, 2012; Santus *et al.*, 2014; The Pham *et al.*, 2015).

One important issue it to determine what type of semantic information can be grasped on the basis of contextual properties, and what part of the meaning of words remains unreachable without complementary knowledge. Recent works focus on this question: Gupta *et al.* (2015) show that referential information is accessible, while results from Herbelot and Ganesalingam (2013) suggest that informativeness (discriminating between more or less contentful words) is difficult to assess on the basis of distributional information. Zarcone *et al.* (2015) show that not only the argument thematic fit to a predicate but also semantic type constraints can be approximated by DSMs to model complement coercion.

In a similar perspective, recent works propose to connect formal and distributional semantics (Guevara, 2011; Grefenstette, 2013), so as to combine the capacity of DSMs to provide semantic representations of word meanings (Erk, 2013; Boleda and Erk, 2015) and the capacity of formal models to account for semantics at the level of complex structures. Compositionality issues have been the focus of many research studies: until recently, most works on DSMs have been concerned with words in isolation, but in the last few years research has been conducted on the extension of these models to process larger semantic units such as phrases and sentences. Two approaches can be considered. The first one consists in taking into account phrases in addition to words as the basic processing units, as did Baldwin *et al.* (2003) in the LSA framework. In this issue, the paper by Périnet and Hamon follows this orientation. Yet this remains a very minority approach, because of the sparsity of data when one con-

siders the distribution of complex units. The second option has generated a large bulk of research. It consists in modeling semantic compositionality within a distributional framework, under the assumption that semantic information about phrases can be computed by combining information about its components. The work by Mitchell and Lapata (2010) proposes a thorough account and evaluation of the combination functions that can be used. Very recently, a task has been proposed on compositional semantics at SemEval (Marelli *et al.*, 2014). Different types of units have been examined, such as Adjective-Noun (Baroni and Zamparelli, 2010),Verb-Noun (Mitchell and Lapata, 2010), sentences. Another area of research concerns the integration of extralinguistic features to complement distributional information with other sources of information, in multimodal models (Bruni *et al.*, 2014).

5. Conclusions and presentation of the papers

Distributional semantics is a young paradigm, but despite its short history we can reliably state that it has been able to gain a large credibility in NLP community and beyond, with increasing interest in cognitive and linguistic research. As shown in this short review, the variety of DSMs is expanding fast, but even more importantly, we have been gaining a much deeper understanding of the effects of their various parameters. The number of semantic tasks that are now addressed by these models has constantly increased, going well beyond the original application of the Distributional Hypothesis to synonym identification. Of course lots of challenges still lie ahead. Under many respects, DSMs still provide a very coarse-grained representation of meaning, and their actual limits and potentialities need to be explored. All this makes distributional semantics a very lively and fascinating research field, as confirmed by the contributions in this special issue.

The four papers published in this issue address a large proportion of the topics we have just listed, such as parameter tuning, evaluation, compositionality or processing of larger lexical units. It is interesting to note that they depart from the dominant trend that consider huge corpora to build distributional models, as three papers out of four are concerned with the treatment of specialized corpora for knowledge acquisition. Two papers are dealing with complex terms, but in a different perspective. Périnet and Hamon ("Analyse distributionnelle appliquée aux textes de spécialité") apply the distributional method to complex terms and propose a solution to normalize the contexts to deal with the problem of sparsity of data. Daille and Hazem ("Méthode semi-compositionnelle pour l'extraction de synonymes des termes complexes") use distributional semantics to generate synonyms of multi-word expressions by leveraging the compositionality properties of these terms. The two other papers focus on the evaluation and improvement of distributional models. Tanguy, Sajous and Hathout's experiment ("Évaluation sur-mesure de modèles distributionnels sur un corpus spécialisé : comparaison des approches par contextes syntaxiques et par fenêtres graphiques") is also based on the treatment of a specialized corpus. They use a specifically designed evaluation dataset to define the best parameters for their distributional model, focus-

ing on the contribution of accurate syntactic information. Ferret's paper ("'Combiner différents critères pour améliorer les thésaurus distributionnels") proposes a way to improve a distributional thesaurus by using a bootstrapping method based on the automatic selection of positive and negative examples of semantic neighbors. The selection procedure takes advantage of the symmetry of the semantic relations, and of the compositionality of compounds.

Acknowledgements

We want to thank the *TAL* journal editors and committee as well as the specific scientific committee. We are particularly grateful to the reviewers for their time and effort to improve this special issue.

Specific scientific committee :

– Marianna Apidianaki – LIMSI, Orsay
– Marco Baroni – CIMeC, Trento
– Ann Bertels – ILT, K.U. Leuven
– Romaric Besançon – CEA, Gif-sur-Yvette
– Yves Bestgen – UCL/CECL, Louvain-La-Neuve
– Gemma Boleda – Université Pompeu Fabra, Barcelone
– Marie Candito – ALPAGE, Paris
– Georgiana Dinu – CIMeC, Trento
– Olivier Ferret – CEA, Gif-sur-Yvette
– Andre Freitas – DERI, National University of Ireland, Galway
– Gregory Grefenstette – INRIA, Saclay
– Thierry Hamon – LIMSI, Paris
– Aurélie Herbelot – Institut für Linguistik, Potsdam
– Mai Ho-Dac – CLLE-ERSS, Toulouse
– Guillaume Jacquet – European Commission, JRC, Ispra
– Olivier Kraif – LIDILEM, Grenoble
– François Morlane-Hondère – LIMSI, Paris
– Yves Peirsman – Leuven
– Laurent Prévot – LPL, Aix-Marseille
– Benoît Sagot – ALPAGE, Paris
– Magnus Sahlgren – Gavagai, Inc., Sweden
– Franck Sajous – CLLE-ERSS, Toulouse
– Sabine Schulte im Walde – IMS, Stuttgart
– Peter Turney – National Research Council Canada, Ottawa
– Tim Van de Cruys – IRIT, Toulouse

6. References

Agirre E., Alfonseca E., Hall K., Kravalova J., Paşca M., Soroa A., "A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches", *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL '09, Association for Computational Linguistics, Stroudsburg, PA, USA, p. 19-27, 2009.

Alfonseca E., Hall K., Hartmann S., "Large-scale computation of distributional similarities for queries", *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, Association for Computational Linguistics, p. 29-32, 2009.

Anguiano E. H., Denis P., "FreDist: Automatic construction of distributional thesauri for French", *Actes de la 18e conférence sur le traitement automatique des langues naturelles – TALN*, p. 119-124, 2011.

Baldwin T., Bannard C., Tanaka T., Widdows D., "An empirical model of multiword expression decomposability", *Proceedings of the ACL 2003 workshop on Multiword expressions: analysis, acquisition and treatment-Volume 18*, Association for Computational Linguistics, p. 89-96, 2003.

Baroni M., Dinu G., Kruszewski G., "Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors", *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, vol. 1, p. 238-247, 2014.

Baroni M., Lenci A., "Distributional memory: A general framework for corpus-based semantics", *Computational Linguistics*, vol. 36, no 4, p. 673-721, 2010.

Baroni M., Lenci A., "How we BLESSed distributional semantic evaluation", *Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics*, Association for Computational Linguistics, p. 1-10, 2011.

Baroni M., Zamparelli R., "Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space", *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, p. 1183-1193, 2010.

Blei D. M., Ng A. Y., Jordan M. I., "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, vol. 3, p. 993-1022, 2003.

Boleda G., Erk K., "Distributional semantic features as semantic primitives – or not", *AAAI Spring Symposium on Knowledge Representation and Reasoning*, Stanford University, USA, 2015.

Bruni E., Tran N.-K., Baroni M., "Multimodal Distributional Semantics.", *Journal of Artificial Intelligence Research (JAIR)*, vol. 49, p. 1-47, 2014.

Budanitsky A., Hirst G., "Evaluating wordnet-based measures of lexical semantic relatedness", *Computational Linguistics*, vol. 32, no 1, p. 13-47, 2006.

Bullinaria J., Levy J. P., "Extracting semantic representations from word co-occurrence statistics: A computational study", *Behavior Research Methods*, vol. 39, p. 510-526, 2007.

Cheung J. C. K., Penn G., "Probabilistic Domain Modelling With Contextualized Distributional Semantic Vectors.", *Association for Computational Linguistics (ACL)*, p. 392-401, 2013.

Claveau V., Kijak E., "Thésaurus distributionnels pour la recherche d'information et vice-versa", *Actes de la 13e Conférence en Recherche d'Information et Applications (CORIA)*, 2015.

Collobert R., Weston J., "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning", *Proceedings of the 25th International Conference on Machine Learning*, Helsinki, Finland, p. 160-167, 2008.

Curran J. R., *From distributional to semantic similarity*, PhD thesis, University of Edinburgh, 2004.

Curran J. R., Moens M., "Improvements in automatic thesaurus extraction", *Proceedings of the ACL-02 workshop on Unsupervised lexical acquisition-Volume 9*, Association for Computational Linguistics, p. 59-66, 2002.

Erk K., "Towards a semantics for distributional representations", *Proceedings of the 10th International Conference on Computational Semantics (IWCS-2013)*, 2013.

Fabre C., Hathout N., Ho-Dac L.-M., Morlane-Hondère F., Muller P., Sajous F., Tanguy L., Van de Cruys T., "Présentation de l'atelier SemDis 2014: sémantique distributionnelle pour la substitution lexicale et l'exploration de corpus spécialisés", *Actes de la conférence Traitement Automatique du Langage Naturel*, Marseille, France, p. 196-205, 2014.

Ferret O., "Identifying Bad Semantic Neighbors for Improving Distributional Thesauri", *51st Annual Meeting of the Association for Computational Linguistics–ACL 2013*, p. 561-571, 2013.

Finkelstein L., Gabrilovich E., Matias Y., Rivlin E., Solan Z., Wolfman G., Ruppin E., "Placing Search in Context: The Concept Revisited.", *ACM Transactions on Information Systems*, vol. 20, n° 1, p. 116-131, 2002.

Grefenstette E., "Towards a formal distributional semantics: Simulating logical calculi with tensors", *Proceedings of the Second Joint Conference on Lexical and Computational Semantics*, Atlanta, USA, p. 1-10, 2013.

Grefenstette G., *Explorations in Automatic Thesaurus Discovery*, Kluwer Academic Publishers, Norwell, MA, USA, 1994.

Guevara E., "Computing semantic compositionality in distributional semantics", *Proceedings of the Ninth International Conference on Computational Semantics*, Association for Computational Linguistics, p. 135-144, 2011.

Gupta A., Boleda G., Baroni M., Padó S., "Mapping conceptual features to referential properties", *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2015.

Habert B., Zweigenbaum P., "Contextual acquisition of information categories", *The Legacy of Zellig Harris: Language and information into the 21st century*, vol. 2, n° 203, p. 139-159, 2002.

Harris Z. S., "Distributional structure", *Word*, vol. 10, n° 2-3, p. 146-162, 1954.

Harris Z. S., *A Theory of Language and Information: A Mathematical Approach*, Clarendon Press, Oxford, 1991.

Herbelot A., Ganesalingam M., "Measuring semantic content in distributional vectors", *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Sofia, Bulgaria, p. 440-445, 2013.

Hill F., Reichart R., Korhonen A., "SimLex-999: Evaluating Semantic Models with (Genuine) Similarity Estimation", *Computational Linguistics*, forthcoming.

Kiela D., Clark S., "A systematic study of semantic vector space model parameters", *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality (CVSC) at EACL*, p. 21-30, 2014.

Koller A., "Top-down questions for distributional semantics", *Presentation at the Workshop on formal and distributional semantics*, Toulouse, 2015.

Landauer T. K., Dumais S. T., "A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge.", *Psychological review*, vol. 104, n° 2, p. 211, 1997.

Lapesa G., Evert S., "A large scale evaluation of distributional semantic models: Parameters, interactions and model selection", *Transactions of the Association for Computational Linguistics*, vol. 2, p. 531-545, 2014.

Lenci A., "Distributional semantics in linguistic and cognitive research", *From context to meaning: Distributional models of the lexicon in linguistics and cognitive science, special issue of the Italian Journal of Linguistics*, vol. 20, n° 1, p. 1-31, 2008.

Lenci A., Benotto G., "Identifying hypernyms in distributional semantic spaces", **SEM 2012: The First Joint Conference on Lexical and Computational Semantics*, Association for Computational Linguistics, Montréal, Canada, p. 75-79, 7-8 June, 2012.

Leviant I., Reichart R., "Judgment Language Matters: Multilingual Vector Space Models for Judgment Language Aware Lexical Semantics", *ArXiv e-prints*, August, 2015.

Levy O., Goldberg Y., "Linguistic Regularities in Sparse and Explicit Word Representations", *Proceedings of the 18th Conference on Computational Natural Language Learning (CoNLL)*, Baltimore, Maryland, USA, p. 171-180, 2014.

Levy O., Goldberg Y., Dagan I., "Improving Distributional Similarity with Lessons Learned from Word Embeddings", *Transactions of the ACL*, vol. 3, p. 211-225, 2015.

Lund C., Burgess K., "Modelling parsing constraints with high-dimensional context space", *Language and cognitive processes*, vol. 12, n° 2-3, p. 177-210, 1997.

Marelli M., Bentivogli L., Baroni M., Bernardi R., Menini S., Zamparelli R., "Semeval-2014 task 1: Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment", *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Dublin, Ireland, p. 1-8, 2014.

McCarthy D., Koeling R., Weeds J., Carroll J., "Unsupervised acquisition of predominant word senses", *Computational Linguistics*, vol. 33, n° 4, p. 553-590, 2007.

McCarthy D., Navigli R., "Semeval-2007 task 10: English lexical substitution task", *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval)*, p. 48-53, 2007.

Mikolov T., Chen K., Corrado G., Dean J., "Efficient Estimation of Word Representations in Vector Space", *In Proceedings of Workshop at ICLR 2013*, p. 1-12, 2013a.

Mikolov T., Yih W.-t., Zweig G., "Linguistic Regularities in Continuous Space Word Representations", *In Proceedings of NAACL-HLT 2013, Atlanta, Georgia*, p. 746-751, 2013b.

Mitchell J., Lapata M., "Composition in Distributional Models of Semantics", *Cognitive Science*, vol. 34, n° 8, p. 1388-1439, 2010.

Morris J., Hirst G., "Non-classical lexical semantic relations", *Proceedings of the HLT-NAACL Workshop on Computational Lexical Semantics*, Association for Computational Linguistics, p. 46-51, 2004.

Muller P., Fabre C., Adam C., "Predicting the relevance of distributional semantic similarity with contextual information", *52nd Annual Meeting of the Association for Computational Linguistics-ACL 2014*, p. 479-488, 2014.

Nazarenko A., Zweigenbaum P., Habert B., Bouaud J., "Corpus-based Extension of a Terminological Semantic Lexicon", *In Recent Advances in Computational Terminology*, John Benjamins, p. 327-351, 2001.

Padó S., Lapata M., "Dependency-based construction of semantic space models", *Computational Linguistics*, vol. 33, n° 2, p. 161-199, 2007.

Padró M., Idiart M., Ramisch C., Villavicencio A., "Nothing like Good Old Frequency: Studying Context Filters for Distributional Thesauri", *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, p. 419-424, 2014.

Peirsman Y., Geeraerts D., "Predicting strong associations on the basis of corpus data", *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, p. 648-656, 2009.

Peirsman Y., Heylen K., Speelman D., "Finding semantically related words in Dutch. Cooccurrences versus syntactic contexts", *Proceedings of the 2007 Workshop on Contextual Information in Semantic Space Models: Beyond Words and Doc- uments*, p. 9-16, 2007.

Sadrzadeh M., Grefenstette E., "A compositional distributional semantics, two concrete constructions, and some experimental evaluations", *Quantum Interaction*, Springer, p. 35-47, 2011.

Sahlgren M., The Word-Space Model, PhD thesis, University of Stockholm, 2006.

Sahlgren M., "The distributional hypothesis", *Italian Journal of Linguistics*, vol. 20, n° 1, p. 33-54, 2008.

Salton G., Wong A., Yang C.-S., "A vector space model for automatic indexing", *Communications of the ACM*, vol. 18, n° 11, p. 613-620, 1975.

Santus E., Lenci A., Lu Q., Schulte im Walde S., "Chasing Hypernyms in Vector Spaces with Entropy", *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, Association for Computational Linguistics, Gothenburg, Sweden, p. 38-42, April, 2014.

Santus E., Yung F., Lenci A., Huang C.-R., "EVALution 1.0: an Evolving Semantic Dataset for Training and Evaluation of Distributional Semantic Models", *Proceedings of the 4th Workshop on Linked Data in Linguistics: Resources and Applications*, Association for Computational Linguistics, Beijing, China, p. 64-69, July, 2015.

The Pham N., Lazaridou A., Baroni M., "A multitask Objective to inject Lexical Contrast into Distributional Semantics", *53rd Annual Meeting of the Association for Computational Linguistics (ACL)*, Beijing, China, p. 21-26, 2015.

Turney P. D., "Similarity of semantic relations", *Computational Linguistics*, vol. 32, n° 3, p. 379-416, 2006.

Turney P. D., "Distributional semantics beyond words: Supervised learning of analogy and paraphrase", *Transactions of the Association for Computational Linguistics (TACL)*, vol. 1, p. 353-366, 2013.

Turney P. D., Pantel P., "From frequency to meaning: Vector space models of semantics", *Journal of artificial intelligence research*, vol. 37, n° 1, p. 141-188, 2010.

Van de Cruys T., "A comparison of bag of words and syntax-based approaches for word categorization", *Proceedings of the ESSLLI Workshop on Distributional Lexical Semantics*, p. 47-54, 2008.

Van de Cruys T., "A non-negative tensor factorization model for selectional preference induction", *Natural Language Engineering*, vol. 16, n° 04, p. 417-437, October, 2010.

Van der Plas L., Tiedemann J., "Finding synonyms using automatic word alignment and measures of distributional similarity", *Proceedings of the COLING/ACL on Main conference poster sessions*, Association for Computational Linguistics, p. 866-873, 2006.

Van der Plas L., Tiedemann J., Manguin J.-L., "Synonym acquisition across domains and languages", *Advances in Distributed Agent-Based Retrieval Tools*, Springer, p. 41-57, 2011.

Zarcone A., Padó S., Lenci A., "Same same but different: Type and typicality in a distributional model of complement coercion", *Proceedings of the NetWordS Final Conference on Word Knowledge and Word Usage*, p. 91-94, 2015.