R. Zamparelli

Similarity, prominence and distributions

"Prominence" is usually understood as an ordering in a discrete scale, such as the one underlying specificity phenomena (e.g. *animate pronouns* > *definites* > *specific indefinites* > *other indefinites* > *predicates*). Departing from this tradition, in this talk I would like to draw the attention of theoretical linguists onto the properties and limits of a set of computational techniques which have received a great deal of attention within the NLP community, but which are potentially of interest also for some areas of theoretical research on language. These 'distributional' techniques are able to give a precise, continuous measure of *semantic distance* between two words (see Mitchell and Lapata (2008)) and more recently, two phrases or sentences (Mitchell and Lapata (2010), Socher et al. (2013)). These methods are based on the statistical properties of the lexical contexts in which a target expression occurs within a huge corpus. The fact that a certain expression, say *dog*, tends to cooccur, on a large scale, with other expressions such as *owner, cat* and *leash* and much less frequently with *funnel* or *hexagonal*, reveals its meaning similarity with other pet-denoting expression, and its distance from other areas of the lexicon. Several studies have shown that these meaning representations, espressed as vectors in a multidimensional space, can be used to model human judgments of word similarity and capture on-line priming effects (see Dumais and Landauer 1997, Turney and Pantel 2010). Moreover, the possibility of compositionally combining lexical vectors to generate distributional representations of larger phrases Baroni et al. (2013) has proved capable of detecting *semantic deviance*, i.e. the fact that (1-a) is judged to be 'less sensible' than (1-b), despite that fact that both describe non-existing objects Vecchi et al. (2011).

(1) a. musical North / salty shadow
b. musical exhaust pipe / salty cotton-candy

There has been some discussion on whether and how these measures can help in some of the 'traditional' tasks of model-theoretical semantics. For instance, McNally (2014) has suggested that the idea that some internal parts of the DP denote kinds, rather than properties, could be implemented in terms of distributions. Similarly, the generation of plausible alternatives in association with focus contexts, a neglected task for semantics, can explain contrasts such as (2).

(2) It is not a dog, it is a{wolf / statue / ??drawer}

On the other hand, current distributional measures are largely insensitive to *negation* and *antonymy* (*big* and *small* tend to be about as similar as *big* and *large*), cannot capture episodic predicates, and are thus neither able not useful to fix pronoun reference. In addition, the most sophisticated methods to assign distributional meanings to sentences (Paperno et al., 2014) typically perform of a par to simply summing up the word vectors of all the words in the sentence. I will argue that this is probably due to a difficulty in finding realistic tests for (distributional) sentence meaning, and I will sketch a view of the integration between distributional and formal semantics in which vectors could replace standard predicative constants within a logic-based framework.

References

Baroni, M., R. Bernardi, and R. Zamparelli (2013). Frege in space: A program for compositional distributional semantics. Submitted, draft available from `http://clic.cimec.unitn.it/composes`.

Dumais, S. and T. Landauer (1997). A solution to Platos problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological review 104*, 211–240.

McNally, L. (2014). Kinds, descriptions of kinds, concepts, and distributions. Paper submitted to a volume from BRIDGE-14, `http://www.upf.edu/pdi/louise-mcnally/_pdf/publications/McNally-kinds-2014.pdf`.

Mitchell, J. and M. Lapata (2008). Vector-based models of semantic composition. In *Proceedings of ACL*, Columbus, OH, pp. 236–244.

Mitchell, J. and M. Lapata (2010). Composition in distributional models of semantics. *Cognitive Science 34*(8), 1388–1429.

Paperno, D., N. Pham, and M. Baroni (2014). A practical and linguistically-motivated approach to compositional distributional semantics. In *Proceedings of ACL 2014 (52nd Annual Meeting of the Association for Computational Linguistics)*, East Stroudsburg PA, pp. 90–99. ACL.

Socher, R., A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts (2013, October). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Washington, USA, pp. 1631–1642. Association for Computational Linguistics.

Turney, P. and P. Pantel (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research 37*, 141–188.

Vecchi, E. M., M. Baroni, and R. Zamparelli (2011). (Linear) maps of the impossible: Capturing semantic anomalies in distributional space. In *Proceedings of the ACL Workshop on Distributional Semantics and Compositionality*, Portland, OR, pp. 1–9.