# Geometry of Meaning from Words to Dialogue Acts


Matthew Purver Mehrnoosh Sadrzadeh

School of Electronic Engineering and Computer Science, Queen Mary University of London

{m.purver,m.sadrzadeh}@qmul.ac.uk


Distributional semantics (DS) can be seen as a geometric model of word employment in context (Firth, 1957), where meanings of words are represented by vectors whose coordinates are obtained from the frequency of word usages in different contexts. Distances between word vectors reflect similarities of word usage, and thus of meaning. The distributions of words and the angles thereof form a basis for a geometry. DS has been applied to word-based language tasks, e.g. to automatic similarity extraction (see e.g. Curran, 2004), disambiguation (see e.g. Schütze, 1998), and entailment (see e.g. Geffet and Dagan, 2005). It has also improved tasks which do not directly deal with word meanings, such as parsing (see e.g. Krishnamurthy and Mitchell, 2013). This poster is motivated by the following programmatic questions:

- It is true that a distributional geometry emerges from the word usage, but does it depend on it? In other words, what (if anything) happens to this geometry if we change the usage of words? What if this is not in the context of the books written in a language, but rather the poems, plays, or quotidian conversations? In such cases, would distributional models provide good representations for word meanings?
- DS seems to have a static nature: vectors are harvested once from a large corpus (which might change every few years, e.g. BNC to ukWaC) and then used and reused in different tasks. The question is, given its static nature, can DS be applied to any language task what-so-ever, even if that task has a dynamic nature?

One can try and answer these questions in their generality. We will, however, focus on one special field where the above mentioned criteria seem to hold: dialogue acts. Here, the main phenomenon under study is that of an utterance. These vary from short elocutionary sounds (sighs, affirmations), to phrases, to half formed sentences, to paragraphs. DS seems to have a potential in modelling dialogue acts. Probabilistic dialogue models that represent distributions over both utterance meanings and contexts, have shown success in practical applications such as human-computer dialogue systems (e.g Young et al., 2013). However, the light that they can shed on linguistic or semantic questions is currently limited, as is the degree to which they can exploit knowledge of fine-grained semantic structure. In order to model meaning in dialogue, we must go beyond individual sentence meanings to look at their role in, and contribution to, a wider structure which emerges through the interaction. Formal models of dialogue tend to account for this dialogue structure in terms of either the relations between utterances (see e.g. Asher and Lascarides, 2003) or their function and effects on some model of context (see e.g. Ginzburg, 2012). In this poster, we suggest a research program that may help us find answers to the original questions raised above and also provide some answers to the questions in the field of dialogue semantics.

Firstly, DS is generated by extracting vector co-ordinates from contexts. The notion of "context" used here tends to be a lexical or syntactic one. Dialogue affords a different notion of syntactic context: not only the role that the word plays in the utterance, but also the role that the utterance plays in the dialogue; the grammatical role of the word in the utterance can be of less importance. The lexical contexts of words in utterances are also different. In distributional semantics, words such as articles and elocutionary expressions are treated as noise and ignored when building word vectors. These expressions become of crucial importance in assigning meaning to utterances and should be taken into account when building a vector space model for words in dialogue. We aim to build and experiment with such *unstructured* vector spaces. Different normalisation schemes and window sizes have been experimented with in DS and on word similarity tasks (e.g. Bullinaria and Levy, 2007), we aim to do the same for dialogue-based spaces and tasks.

Secondly, recent advances in DS extend them from words to sentences, enabling researchers to reason about the geometry of sentence representations. This is despite the fact that sentence representations are not so much about distributions of sentences in contexts, but rather about the effect of grammatical constructions on representations of words. To apply these models to dialogue acts, we face two main choices.

1. One can ignore grammatical structure, and use compositional operators that act only over sets or sequences of words. Example of such operations are addition and point wise multiplication, which in some tasks have performed a par with more complicated models. These can then be directly applied to model utterance

semantics, as has been done for tagging experiments (e.g. Kalchbrenner and Blunsom, 2013; Milajevs et al., 2014). It remains to see how such models will perform in other dialogue tasks, such as paraphrasing.

2. The starting point of tensor-based compositional models (e.g. Coecke et al., 2010) is grammatical structure. They use the POS-tag information provided by a type-logic and provide us with a way to articulate predicate-argument structure in vector spaces, e.g. viewing verbs as distributional relations over their arguments. We aim to develop a similar model for dialogue semantics. Here there are also two main choices.

   - Despite often being fragmentary, dialogue acts have an utterance-internal grammatical structure of their own, and attempts have been made to express this using dynamic grammars (see e.g. Kempson et al., 2001), and incremental parsers (e.g. Purver et al., 2011). Dynamic grammars have type-logical formalizations which can be used to develop a functorial passage in the style of Coecke et al. (2010). We aim to use these to homomorphically transfer the grammatical structures of dialogue acts to linear maps in vector spaces and define a tensor-based semantic model for dialogues.
   - Dialogues themselves have a structure and semantics which emerge from the context update functions of utterances, which can themselves be based on context updates of words and phrases. These updates can be seen in terms of association between utterances and distributions over the contexts that they relate and the responses that they create, and perhaps the same is true for words and phrases (see e.g. Purver and Ginzburg, 2004). We aim to use the update models, instead of or together with dynamic grammars, as a starting point for a compositional distributional model of dialogue.

Finally, one can ask what can dialogue semantics do for distributional models? The notion of context in the latter is rather limited and this can lead to surprising results such as an apparent similarity in meaning between terms like "happy" and "sad", or "hello" and "goodbye". Dialogue provides us with a different notion of context – how can this be exploited to learn word meanings from distributions, and what difference does this make?

# References


Asher, N. and A. Lascarides (2003). *Logics of Conversation*. Cambridge University Press.

Bullinaria, J. A. and J. P. Levy (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, 510–526.

Coecke, B., M. Sadrzadeh, and S. Clark (2010). Mathematical foundations for a compositional distributional model of meaning. *Linguistic Analysis 36*(1–4), 345–384. Special Issue: A Festchrift for Joachim Lambek.

Curran, J. (2004). *From Distributional to Semantic Similarity*. Ph. D. thesis, School of Informatics, University of Edinburgh.

Firth, J. R. (1957). A Synopsis of Linguistic Theory, 1930-1955. *Studies in Linguistic Analysis*, 1–32.

Geffet, M. and I. Dagan (2005). The distributional inclusion hypotheses and lexical entailment. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, Stroudsburg, PA, USA, pp. 107–114.

Ginzburg, J. (2012). *The Interactive Stance: Meaning for Conversation*. Oxford University Press.

Kalchbrenner, N. and P. Blunsom (2013). Recurrent convolutional neural networks for discourse compositionality. In *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*, Sofia, Bulgaria, pp. 119–126.

Kempson, R., W. Meyer-Viol, and D. Gabbay (2001). *Dynamic Syntax: The Flow of Language Understanding*. Blackwell.

Krishnamurthy, J. and T. Mitchell (2013). *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*, Chapter Vector Space Semantic Parsing: A Framework for Compositional Vector Space Models, pp. 1–10. Association for Computational Linguistics.

Milajevs, D., D. Kartsaklis, M. Sadrzadeh, and M. Purver (2014). Evaluating neural word representations in tensor-based compositional settings. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, pp. 708–719.

Purver, M., A. Eshghi, and J. Hough (2011). Incremental semantic construction in a dialogue system. In J. Bos and S. Pulman (Eds.), *Proceedings of the 9th International Conference on Computational Semantics*, Oxford, UK, pp. 365–369.

Purver, M. and J. Ginzburg (2004). Clarifying noun phrase semantics. *Journal of Semantics 21*(3), 283–339.

Schütze, H. (1998). Automatic word sense discrimination. *Computational Linguistics 24*(1), 97–123.

Young, S., M. Gašić, B. Thomson, and J. D. Williams (2013). POMDP-based statistical spoken dialog systems: A review. *Proceedings of the IEEE 101*(5), 1160–1179.