A practical grounded compositional semantics for situated interaction

David Schlangen, Bielefeld University. `david.schlangen@uni-bielefeld.de`

Our starting point is our interest in situated human/machine interaction. To be useful in situated settings, an agent must be capable of referring to things currently present both for it and its interlocutor (things it "sees"), but also of referring to things it may have seen earlier but which aren't currently physically present (for which the discourse provides the referent).

For this, the agent needs a conceptual system that underwrites its use of language. We assume that this conceptual system is structured, with concepts encapsulating knowledge of different types: e.g., where appropriate, knowledge about the appearance of instances of the category the concept represents, enabling reference (resolve "the box" to this object); but also knowledge about how a given category is related to other categories, allowing for inferences ("there is a red box", hence "there is a coloured container"). The concepts are learned in interaction in different types of situations: ostensive definition ("this is a glove"), explicit definition ("it's called a mitten if it doesn't have fingers and a glove if it has fingers"), and implicit definition (occurence in context, e.g. "she was wearing a scarf and mittens, as it was a cold night.").

It might seem that there is a direct connection between learning situations and types of knowledge (with e.g. ostensive definition providing training data for representations of appearances), but we assume that there is "cross-fertilisation", enabling what in the psycholinguistic literature is called "fast-mapping" and in the machine learning literature "few shot" / "zero shot" learning. (If I know that a "wampimuk" is a small rodent but have never seen one, I can resolve, with high probability, a reference using that term, if I see a small rodent for which I lack another specific label.) Both the learning process as well as the application process are fallible, and so we assume that the conceptual system must be open for explicit interactive negotiation ("meta-semantic interaction").

I will outline our approach to grounding lexical semantics and contrast it to related work, making the point that most other work focusses only on one aspect or the other – work in computational linguistics for example on learning from verbal context, work in computer vision on learning from ostensive definition / labelled instances – whereas we see the combination and synergy of aspects as crucial. I will describe our work on modelling exophoric reference using neural networks taking images as input (e.g., Schlangen et al. 2016), our experiments on cross-fertilisation between distributional knowledge and referential / visual type of knowledge (e.g., Zarrieß & Schlangen, 2017), our initial work on modelling meta-semantic interaction (Schlangen 2016), and our plans for future work exploring the interconnections between the different learning situations, types of knowledge on the one hand, and the conceptual system and the compositional system on the other.

References

- Schlangen D. Grounding, Justification, Adaptation: Towards Machines That Mean What They Say. In: Proceedings of the 20th Workshop on the Semantics and Pragmatics of Dialogue (JerSem). 2016. PDF
- Schlangen D, Zarrieß S, Kennington C. Resolving References to Objects in Photographs using the Words-As-Classifiers Model. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016). Berlin; 2016. PDF
- Zarrieß S, Schlangen D. Obtaining referential word meanings from visual and distributional information: Experiments on object naming. In: Proceedings of 55th annual meeting of the Association for Computational Linguistics (ACL). Vancouver. 2017 PDF
- Zarrieß S, Schlangen D. Is this a Child, a Girl, or a Car? Exploring the Contribution of Distributional Similarity to Learning Referential Word Meanings. In: Short Papers – Proceedings of the Annual Meeting of the European Chapter of the Association for Computational Linguistics (EACL). 2017 PDF