Annotation and automatic classification of situation entity types

Annemarie Friedrich & Alexis Palmer

Universität des Saarlandes

Linguistic expressions form patterns in discourse. One way to analyze text passages is in terms of the situation entities (SEs) that they introduce to the discourse. In her work on modes of discourse, Smith (2003) distinguishes several SE types, which are expressed at the clause level. STATES introduce specific properties of specific individuals to the discourse (*Carl is a cat*), EVENTS introduce a specific event (*Carl entered the room*), GENERALIZING SENTENCES report regularities related to specific individuals (*Mary often feeds my cats*) and GENERIC SENTENCES make statements about kinds (*Cats are popular*). There are also ABSTRACT ENTITIES, which are clausal complements of verbs of knowledge or belief (*I know/believe that Mary likes cats*).

Although these categories are clearly distinct from one another on theoretical grounds, in practice it can be difficult to cleanly draw boundaries between them. Our research project¹ aims to assess the applicability of the classification of types as described by Smith (2003): to what extent can situations be classified easily, which borderline cases occur, and how do humans perform on this task?

We annotate a corpus of clauses (from $MASC^2$ and Wikipedia) with their SE type, as well as information about several features which we found helpful for distinguishing SE types (Friedrich & Palmer, 2014b). Clauses can make a statement about a particular individual or about a class or kind; we call this feature the genericity of the main referent. The verb's fundamental aspectual class in context, stative or dynamic, is marked. Finally, we mark the clauses' habituality, which expresses whether the clause makes a statement about some regularity or about an episodic event.

We develop and evaluate automatic systems classifying SEs, and sub-tasks which have (partially) been studied by the computational linguistics community, but for which no large annotated corpora are available. For example, we have worked on automatically predicting whether a verb is used with a stative or dynamic meaning (Friedrich & Palmer, 2014a). Our next step focuses on the genericity of clauses and noun phrases, aiming at automatically detecting whether a clause makes a statement about a kind or class. In addition, we aim to improve prior work (Palmer et al., 2007) on automatically classifying clauses for SE types.

References

- Friedrich, Annemarie & Alexis Palmer. 2014a. Automatic prediction of aspectual class of verbs in context. In *Proceedings of ACL 2014*, Baltimore, USA.
- Friedrich, Annemarie & Alexis Palmer. 2014b. Situation entity annotation. In *Proceedings* of the 8th Linguistic Annotation Workshop (LAW VIII), Dublin, Ireland.
- Palmer, Alexis, Elias Ponvert, Jason Baldridge & Carlota Smith. 2007. A sequence model for situation entity classification. In *Proceedings of ACL 2007*, Prague, Czech Republic.
- Smith, Carlota S. 2003. *Modes of discourse: The local structure of texts.* Cambridge University Press Cambridge.

¹http://sitent.coli.uni-saarland.de

²http://www.americannationalcorpus.org/MASC/About.html