Preposition Supersenses in German–English Parallel Data

Jakob Prange
Department of Computer Science
Georgetown University
`jp1724@georgetown.edu`

Nathan Schneider
Department of Computer Science
Department of Linguistics
Georgetown University
`Nathan.Schneider@georgetown.edu`

1 Introduction

Despite being often disregarded as merely grammatical function words, prepositions have the power to express a great variety of semantic relations (Tyler and Evans, 2003; Durrell and Brée, 1993; Müller et al., 2010). Knowing which relation is present in a certain context could greatly enhance the performance of machine translation (Hashemi and Hwa, 2014) and information extraction systems.

This work presents a pilot study in annotating German prepositions with supersenses from the SNACS hierarchy (Schneider et al., 2018) and aligning them with parallel English data. We thereby provide a perspective of how this kind of semantic analysis can benefit multilingual learning tasks, but also identify gaps in the current schema.

While we generally take into account the syntactic and semantic interactions between prepositions and case in German (Zwarts, 2005; Van Riemsdijk, 2007), the SNACS framework relies on lexical items as markables and does not support the annotation of morphological case markers. Following Blodgett and Schneider (2018), we include possessive pronouns, but exclude genitive case marking on nouns, which in German is realized inflectionally, rather than through clitics.

We also incorporate the *construal analysis* introduced by Hwang et al. (2017): Each preposition token is labeled with (1) the *scene role* occupied by the PP in the specific context, which we assume to be largely language-independent and often preserved in translation; and (2) the *function* prototypically contributed by the preposition type, which may be directly dependent on the lexical choice, a verb's preferred argument structure, or other language-specific parameters.

2 Annotation Study

One linguistically trained annotator identified and labeled all prepositional expressions in chapters 1 and 4 of the German translation of Saint-Exupéry's *Le Petit Prince*. These were then manually aligned with the SNACS-annotated parallel English text (Schneider et al., 2018). The passage contained 1,361 tokens in German and 1,672 tokens in English.

Results. Even though the proportion of preposition tokens is similar in both languages (109=8% in German vs. 168=10% in English), only 73 can be aligned to an equivalent lexical item. Out of these, almost 80% are labeled with the same scene role in both languages, while only 55% share the same function. This highlights the importance of the construal analysis in a multilingual context.

However, finding these lexical counterparts is difficult for several reasons: In many situations where English requires prepositions or clitics to express certain kinds of semantic roles (e.g., *possessor*, *recipient*), German does not, and in fact prefers the use of genitive or dative case inflections, which are excluded from the annotation. Additionally, the fact that both texts are translations from a third language, French, leads to diverging syntactic realizations.

3 Future Work

We argue that preposition supersenses can help to exploit semantic similarities in German and English that would otherwise be overshadowed by language-specific construals and constructions. To increase coverage for languages morphologically richer than English, we will investigate how non-lexical case markers can be added to the current framework, and use that to build a system that jointly learns to assign supersenses in a multilingual environment.

References

Austin Blodgett and Nathan Schneider. 2018. Semantic supersenses for English possessives. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, Miyazaki, Japan. ELRA.

Martin Durrell and David Brée. 1993. German temporal prepositions from an english perspective. *The Semantics of Prepositions: From Mental Processing to Natural Language Processing*, 3:295.

Homa B Hashemi and Rebecca Hwa. 2014. A comparison of MT errors and ESL errors. In *LREC*, pages 2696–2700.

Jena D Hwang, Archna Bhatia, Na-Rae Han, Tim O'Gorman, Vivek Srikumar, and Nathan Schneider. 2017. Double trouble: The problem of construal in semantic annotation of adpositions. In *Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (* SEM 2017)*, pages 178–188.

Antje Müller, Olaf Hülscher, Claudia Roch, Katja Keßelmeier, Tobias Stadtfeld, Jan Strunk, and Tibor Kiss. 2010. An annotation schema for preposition senses in German. In *Proceedings of the Fourth Linguistic Annotation Workshop*, pages 177–181. Association for Computational Linguistics.

Nathan Schneider, Jena D Hwang, Vivek Srikumar, Jakob Prange, Austin Blodgett, Sarah R Moeller, Aviram Stern, Adi Bitan, and Omri Abend. 2018. Comprehensive supersense disambiguation of English prepositions and possessives. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia. ACL. To appear.

Andrea Tyler and Vyvyan Evans. 2003. *The semantics of English prepositions: Spatial scenes, embodied meaning, and cognition*. Cambridge University Press.

Henk Van Riemsdijk. 2007. Case in spatial adpositional phrases: The dative-accusative alternation in German. *Pitar Mos: A building with a view. Festschrift for Alexandra Cornilescu*, pages 1–23.

Joost Zwarts. 2005. The case of prepositions: Government and compositionality in German PPs. In *21st Annual Meeting of the Israel Association for Theoretical Linguistics, Haifa*.