

How to analyze semantic roles across interfaces: taking FrameNet to discourse level.

Levi Remijnse¹, Pia Sommerauer,² Antske Fokkens³ & Piek Vossen⁴
^{1,2,3,4}Vrije Universiteit at Amsterdam, 1.remijnse@vu.nl, 2.pia.sommerauer@vu.nl,
3.antske.fokkens@vu.nl, 4.p.t.j.m.vossen@vu.nl

Keywords: FrameNet, discourse, semantic roles, referential grounding

Rather than investigating the conceptual schemas evoked by language, this paper takes a referentially grounded corpus annotated in the Dutch FrameNet (DFN) annotation tool (Remijnse et al., 2022, Postma et al., 2020) and starts from conceptual schemas to investigate framing in language. By taking events in the world as the starting point of compilation, the corpus shows large variation of texts referencing single events. The ultimate goal is to annotate the network of frames and see how they form a semantic structure at both text and corpus level. This way, the frames distribute semantic roles on different linguistic interfaces. So far, the discussion surrounding semantic roles has largely focused on the semantics-syntax interface (Fillmore, 1968; Dowty 1991; Primus, 2016), resulting in many annotated corpora, such as PropBank (Palmer et al., 2005), FrameNet (Ruppenhofer et al., 2016) and VerbNet (Schuler, 2005), and has developed into the task of Semantic Role Labeling (SRL) (Gildea & Jurafsky, 2002). However, the sentences below display alternative ways of realizing roles in addition to syntax.

- (1) a. William had to transfer between a lot of flights during his trip to Denmark.
b. Yesterday, he finally arrived.

- (2) The thief ordered a cup of coffee.

In the sequence in (1), the destination role of the verb *arrive* in (1b) is realized but sentence-externally in (1a) by *Denmark*. Although there has been some attention to discourse-licensed roles in cognitive science (Carlson & Tanenhaus, 1988) and SRL (Gerber & Chai, 2012; Ruppenhofer et al., 2010), this phenomenon has largely been ignored in classic literature and also in annotated corpora.

In addition to discourse, semantic roles can also remain completely unexpressed. Sequence (1) does not mention any source role. This role is then not licensed linguistically, but pragmatically inferred. Furthermore, *thief* in (2) is the agent of order, but at word level also incorporates an agentive role with regards to the concept of stealing. As far as we know, these types of semantic roles have gained little to no attention in literature and little is known about how it connects to other expressions in the discourse. Fillmore (1986) and Ruppenhofer et al. (2016) propose a variety of Null-Instantiations, but since the FrameNet paradigm is applied at syntax level, those unexpressed roles are lacking extra-sentential antecedents in annotation.

All of these semantic role realizations are captured in our annotations. Text-mentions are both linked to structured data and annotated with FrameNet frames. Referencing the same events, the texts together form a coherent story but differ in their narrative structure and framing. In order to capture the dynamic narratives of each text, the FrameNet annotations are adapted to the discourse level, i.e., within a text, frame elements are annotated across sentence boundaries, in addition to word level, and as implicatures when not found in the document. This way, we cover the realization of frame elements across all aforementioned linguistic interfaces and provide an annotation scheme in which we can observe how - in natural language data - these interfaces are utilized to distribute the roles.

The links to structured data are then used to investigate this distribution as a function of reference in combination with pragmatic factors. We expect, for example, that topicalization affects foregrounding and backgrounding principles. Focalized participants are foregrounded and therefore overtly express a high number of roles in syntax, whereas non-focalized participants are backgrounded by minimal expression, which suffices as the antecedent of roles across sentences. Thus, the realization of a role in natural language is subjected to topicalization of its referent.

References

- Carlson, Greg. N. & Tanenhaus, Michael. K. Tanenhaus. 1988. Thematic roles and language comprehension. *Syntax and semantics*, 21, 263-288.
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language*, 67(3), 547-619.

- Fillmore, Charles. J. 1968. Lexical entries for verbs. *Foundations of language*, 373-393.
- Fillmore, C. J. 1986. Pragmatically controlled zero anaphora. In *Annual Meeting of the Berkeley Linguistics Society*, 12, 95-107.
- Gerber, Matthew & Joyce. Y. Chai. 2012. Semantic role labeling of implicit arguments for nominal predicates. *Computational Linguistics*, 38(4), 755-798.
- Gildea, Daniel & Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational linguistics*, 28(3), 245-288.
- Kingsbury, Paul R. & Martha Palmer. 2002. From treebank to propbank. In *LREC*. 1989-1993.
- Postma, Marten, Levi Remijnse, Filip Ilievski, Antske Fokkens, Sam Titarsolej, & Piek Vossen. 2020. Combining Conceptual and Referential Annotation to Study Variation in Framing. In *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*. 31-40.
- Primus, Beatrice. 2016. Participant roles. In: Nick Riemer (ed.): *The Routledge Handbook of Semantics*. London. 403–418.
- Remijnse, Levi, Piek Vossen, Antske Fokkens & Sam Titarsolej. 2022. Introducing Frege to Fillmore: A FrameNet Dataset that Captures both Sense and Reference. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. 39-50.
- Ruppenhofer, Josef, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson & Jan Scheffczyk. 2016. *FrameNet II: Extended theory and practice*. International Computer Science Institute.
- Ruppenhofer, Josef, Caroline Sporleder, Roser Morante, Colin Baker, & Martha Palmer. 2010. Semeval-2010 task 10: Linking events and their participants in discourse. In *Proceedings of the 5th International Workshop on Semantic Evaluation*. 45-50.
- Schuler, Karin Kipper. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. University of Pennsylvania.