

Annotating Semantic Frames in RRGparbank

Tatiana Bladier, Kilian Evang, Laura Kallmeyer
Heinrich Heine University Düsseldorf
first.last@hhu.de

Introduction. This paper describes the ongoing frame-semantic annotation of RRGparbank[3], a multilingual parallel treebank for Role and Reference Grammar (RRG) that contains syntactic RRG annotations of George Orwell’s novel *1984* and its translations. RRGparbank is one of the first resources for which RRG has been applied to large amounts of real-world data. Annotating RRGparbank with semantics creates new possibilities of data-driven NLP applications based on RRG, for example semantic parsers for various languages. The paper gives an overview of the decisions for annotating semantics of four languages in RRGparbank (English, German, French, and Russian). We also describe our annotation decisions for some selected linguistic phenomena. As a proof of concept for possible applications, we train a multilingual semantic parser on the treebank data.

Semantic annotation decisions. Semantic frames and roles are important for capturing semantic relations between the event participants and also between the sentences (for example to use for automated semantic reasoning). We follow the VerbAtlas frame lexicon [4] for annotation of semantic frames and roles in RRGparbank. We chose VerbAtlas because it uses a relatively small set of roles, which is good for training NLP models, but sufficiently large to represent different facets of meaning. Frames are also explicitly linked to English verbal word senses, covering all verbal synsets in WordNet [6], which is in turn linked to verb senses in other languages. Finally, the VerbAtlas roleset is to a large extent compatible with that of VerbNet [7], which is itself organized into a hierarchy from coarse-grained to fine-grained roles, which makes it possible to train models of differing granularity, depending on the application. Each sentence in RRGparbank is annotated independently by two linguistically trained annotators and afterwards adjudicated by a semantic expert.

We annotate only heads of semantic role spans and not full spans—for example, if the NP *the swift answer* fills a Theme role, the word *answer* is annotated as Theme. The full spans of semantic roles can be reconstructed deterministically from the corresponding syntactic trees. Figure 1 shows an example of the annotation interface. The rows represent semantic frames along with the annotated semantic roles. The frame labels are given in all capital letters. Each sentence is annotated independently by two linguistically trained annotators and afterwards adjudicated by a semantic expert.

Selected phenomena. We define *verbal idioms* as verbs with fixed figurative meanings that have one or more fixed arguments. We annotate the frame and roles for non-fixed arguments based on the figurative meaning, and fixed arguments with the pseudo-role VID.

(1) Verbal idioms

- a. Don’t spill_[SPEAK] the beans_[VID]
- b. Are you_[Agent] pulling_[JOKE] my_[Theme] leg_[VID]?

If a verb is used figuratively outside of a fixed idiom and no synset explicitly covers that figurative use, we choose a frame and roles based on the literal meaning.

(2) Figurative Use of Predicates

- a. Die Augen_[Agent] bohrten_[HOLE_PIERCE] sich_[Reflexive] tief_[Extent]
The eyes pierced themselves deeply
in Winstons Blick_[Destination]
in Winstons’s eyes
‘the eyes looked deep into Winston’s own’

We annotate light verbs with the predicate that the dependent expresses, and mark the dependent with the special role LVC.

15th International Conference on Computational Semantics IWCS-15. Ed. by (accepted). Nancy, 2023. [3] Tatiana Bladier et al. “RRGparbank: A Parallel Role and Reference Grammar Treebank”. In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, June 2022, pp. 4833–4841. URL: <https://aclanthology.org/2022.lrec-1.517>. [4] Andrea Di Fabio, Simone Conia, and Roberto Navigli. “VerbAtlas: a Novel Large-Scale Verbal Semantic Resource and Its Application to Semantic Role Labeling”. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 627–637. DOI: 10.18653/v1/D19-1058. URL: <https://aclanthology.org/D19-1058>. [5] Kilian Evang et al. “Improving Low-resource RRG Parsing with Cross-lingual Self-training”. In: *Proceedings of the 29th International Conference on Computational Linguistics*. Gyeongju, Republic of Korea: International Committee on Computational Linguistics, Oct. 2022, pp. 4360–4371. URL: <https://aclanthology.org/2022.coling-1.384>. [6] Christiane D. Fellbaum. “WordNet : an electronic lexical database”. In: *Language* 76 (2000), p. 706. [7] Karin Kipper Schuler. “VerbNet: A broad-coverage, comprehensive verb lexicon”. PhD thesis. University of Pennsylvania, 2005.