

Getting the Roles Right: Using FrameNet in NLP

Collin F. Baker, Nathan Schneider, Miriam R. L. Petruck, Michael Ellsworth

The FrameNet lexical database (Fillmore & Baker 2010; Ruppenhofer *et al.* 2006) <http://framenet.icsi.berkeley.edu>), covers roughly 13,000 lexical units (word senses) for the core English lexicon, associating them with roughly 1,200 fully defined semantic frames; these frames and their roles cover the majority of event types in everyday, non-specialist text, and they are documented with 200,000 manually annotated examples. This tutorial will teach attendees what they need to know to start using the FrameNet lexical database as part of an NLP system. We will cover the basics of Frame Semantics, explain how the database was created, introduce the Python API and the state of the art in automatic frame semantic role labeling systems; and we will discuss FrameNet collaboration with commercial partners. Time permitting, we will present new research on frames and annotation of locative relations, as well as corresponding metaphorical uses, along with information about how frame semantic roles can aid the interpretation of metaphors.

- Introduction
 - FrameNet and its relevance to NLP
 - crucial differences from other resources
 - * WordNet
 - * PropBank
 - * AMR
 - FrameNets in other languages
 - * Spanish FN
 - * Swedish FN++
 - * Japanese FN
 - * Multilingual FrameNet
- The Components of Berkeley FrameNet
 - Frames
 - Frame Elements (roles)
 - Frame-to-frame relations
 - Lexicographic annotation
 - Full-text annotation
- Demo of the FrameNet website
- Using the Python API and NLTK integration
- How FrameNet annotation works
 - Vanguarding, subcorporation, and annotation
 - Frame creation
 - Current research on procedural improvements (crowdsourcing, etc.).

- Overview of ASRL research (including SEMAFOR)
- Applications of FrameNet/ASRL
 - FN Brasil: World Cup, Olympics
 - DAC collaboration
- Q&A / Discussion

Collin Baker (International Computer Science Institute, collinb@icsi.berkeley.edu), has been Project Manager of the FrameNet Project since 2000. His research interests include FrameNets in other languages (Lönneker-Rodman & Baker 2009), aligning FrameNet to other lexical resources (Fellbaum & Baker 2013; Ferrández *et al.* 2010a), linking to ontologies and reasoning (Scheffczyk *et al.* 2010), and the frame semantics of metaphor.

Nathan Schneider (University of Edinburgh, nschneid@inf.ed.ac.uk, <http://nathan.c1>) has worked on a coarse-grained representation for lexical semantics (2014 dissertation at Carnegie Mellon University) and the design of the Abstract Meaning Representation (AMR; Banarescu *et al.* 2014). Nathan helped develop the leading open-source frame-semantic parser for English, SEMAFOR (Das *et al.* 2010, 2014) (<http://demo.ark.cs.cmu.edu/parse>), as well as a Python interface to the FrameNet lexicon (with Chuck Wooters) that is part of the NLTK suite.

Miriam R. L. Petruck (International Computer Science Institute, miriamp@icsi.berkeley.edu) received her PhD in Linguistics from the University of California, Berkeley. A key member of the team developing FrameNet almost since the project's founding, her research interests include semantics, knowledge base development, grammar and lexis, lexical semantics, Frame Semantics and Construction Grammar.

Michael Ellsworth (International Computer Science Institute, infinity@icsi.berkeley.edu) has been involved with FrameNet for well over a decade. His chief focus is on semantic relations in FrameNet (Ruppenhofer *et al.* 2006), how they can be used for paraphrase (Ellsworth & Janin 2007), and mapping to other resources (Scheffczyk & Ellsworth 2006; Ferrández *et al.* 2010b). Increasingly, he has examined the connection of FrameNet to syntax and the Constructicon (Torrent & Ellsworth 2013; Ziem & Ellsworth to appear), including in his pending dissertation on the constructions and frame semantics of emotion.

References

- ELLSWORTH, MICHAEL, & ADAM JANIN. 2007. Mutaphrase: Paraphrasing with framenet. In *Proceedings of the Workshop on Textual Entailment and Paraphrasing*, Prague. Association for Computational Linguistics.
- FELLBAUM, CHRISTIANE, & COLLIN BAKER. 2013. Comparing and harmonizing different verb classifications in light of a semantic annotation task. *Linguistics* 51.707–728.
- FERRÁNDEZ, ÓSCAR, MICHAEL ELLSWORTH, RAFAEL MUÑOZ, & COLLIN F. BAKER. 2010a. Aligning FrameNet and WordNet based on semantic neighborhoods. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, ed. by Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odjik, Stelios Piperidis, Mike Rosner, & Daniel Tapias, 310–314, Valletta, Malta. European Language Resources Association (ELRA).
- FERRÁNDEZ, OSCAR, MICHAEL ELLSWORTH, RAFAEL MUÑOZ, & COLLIN F. BAKER. 2010b. A graph-based measure of FrameNet-WordNet alignment. In *Proceedings of ICGL 2010, Second International Conference on Global Interoperability for Language Resources*, Hong Kong.
- FILLMORE, CHARLES J., & COLLIN F. BAKER. 2010. A frames approach to semantic analysis. In *Oxford Handbook of Linguistic Analysis*, ed. by Bernd Heine & Heiko Narrog, 313–341. OUP.

- LÖNNEKER-RODMAN, BIRTE, & COLLIN F. BAKER. 2009. The FrameNet model and its applications. *Natural Language Engineering* 15.415–453.
- RUPPENHOFER, JOSEF, MICHAEL ELLSWORTH, MIRIAM R. L. PETRUCK, CHRISTOPHER R. JOHNSON, & JAN SCHEFFCZYK. 2006. *FrameNet II: Extended Theory and Practice*. Berkeley, California: International Computer Science Institute. Distributed with the FrameNet data.
- SCHEFFCZYK, JAN, COLLIN F. BAKER, & SRINI NARAYANAN. 2010. Reasoning over natural language text by means of FrameNet and ontologies. In *Ontology and the Lexicon: A Natural Language Processing Perspective*, ed. by Chu-Ren Huang, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci, Alessandro Oltramari, & Laurent Prévot, Studies in Natural Language Processing, chapter 4, 53–71. Cambridge, UK: Cambridge University Press. Expanded version of paper at OntoLex, 2006. (ISBN-13: 9780521886598).
- , & MICHAEL ELLSWORTH. 2006. Improving the quality of framenet. In *Proceedings of the Workshop on Quality assurance and quality measurement for language and speech resources*, ed. by Steven Krauwer & Uwe Quasthoff, 8–13, Genoa, Italy. LREC.
- TORRENT, TIAGO, & MICHAEL ELLSWORTH. 2013. Behind the labels: Criteria for defining analytical categories in FrameNet Brasil. *Veredas* 17.44–65.
- ZIEM, ALEXANDER, & MICHAEL ELLSWORTH. to appear. Exklamativsätze im framenet-konstruktikon am beispiel des englischen. In *Satztypen und Constructionen im Deutschen*, ed. by Rita Finkbeiner & Jörg Meibauer.