Frame Semantics across Languages: Towards a Multilingual FrameNet

Collin F. Baker[1]   **Michael Ellsworth**[1]   **Miriam R. L. Petruck**[1] **Swabha Swayamdipta**[2]

1. International Computer Science Institute, 1947 Center St. Suite 600, Berkeley, California, 94704
2. Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, 15213
{collinb,infinity,miriamp}@icsi.berkeley.edu, swabha@cs.cmu.edu

Abstract

FrameNet is a lexical resource that provides rich semantic representations of the core English vocabulary based on Fillmore's Frame Semantics, with more than 200k manually annotated examples. Resources based on FrameNet have now been created for roughly a dozen languages. This workshop will present current research on aligning Frame Semantic resources across languages and automatic frame semantic parsing in English and other languages. We will explore the extent to which semantic frames are similar across languages and the implications for theories of semantic universals, the practice of translation (whether human or machine), and multilingual knowledge representation. Does not require prior familiarity with Frame Semantics.

1 Description

The FrameNet Project at the International Computer Science Institute (ICSI, http://framenet.icsi.berkeley.edu) has created a detailed lexical database of contemporary English, (currently more than 13,000 lexical units in 1,200 semantic frames) based on Frame Semantics (Fillmore (1977), (1985)). NLP researchers have shown that such representations are useful in diverse applications such as question answering, text-to-scene systems, dialog systems, and social network extraction. Separate research projects have now developed Frame Semantic lexical databases for roughly a dozen languages (including Brazilian Portuguese, Chinese, Hebrew, Japanese, Korean, and Swedish) based in varying degrees on the original FrameNet structure, methodology and annotation practices (Ruppenhofer et al., 2016). Most have taken the ICSI English frames as a starting point and have found that the majority of target-language words fit comfortably in those frames.

The FrameNet team is developing alignments across these FrameNets, seeking to understand cross-linguistic similarities and differences in framing. Going beyond alignments between frames, we also use techniques such as multilingual word vectors (Hermann and Blunsom, 2014) to cluster and align lexical units (each a single sense of a word in a frame) across languages. The underlying research questions include: To what extent are semantic frames the same cross-culturally and cross-linguistically? Where they differ, what are the reasons for these differences? Will translations to be "frame preserving"?

This tutorial will discuss the methodology and status of the alignment effort, and of a recently launched parallel manual annotation task, and theoretical issues that have emerged in this area of research, including the interplay between semantic frames and constructions in different languages. We will also report on the state of the art in automatic frame semantic role labeling for English (Swayamdipta et al., 2017) and for other languages. This can be regarded as a structured prediction task which maps a sentence to a graph with nodes for each predicator and its arguments and adjuncts, linked by arcs representing the frame semantic roles. Recent approaches rely on neural architectures to learn representations which enforce global consistency for each classification decision and can learn from disparate data.

Participants in this tutorial will learn about:

1. Multilingual FrameNet, its methodology and practices

2. Cross-linguistic similarities and differences among the languages

3. Principles and challenges of multilingual lexical alignment

4. Issues of representation in frame vs. constructional analyses

5. Recent advances in Frame Semantic parsing (a.k.a. Frame Semantic role labeling)

6. Potential Applications of multilingual FrameNet databases

2   Outline of Tutorial

- Frame Semantics and FrameNet (Miriam R. L. Petruck)
    - Frames, Frame Elements (roles), and Lexical Units
    - Annotation and Reports
    - Frame Relations and Frame Element Relations
    - Construction Grammar and Constructicons
    - Constructions implicit in FrameNet Annotation
- The Multilingual FrameNet Project (Collin Baker)
    - FrameNets in Languages other than English
    - Variations in Sources, Goals, and Practices
    - Interactions between Frames and Constructions
    - Alignment across Languages
    - Alignment Algorithms and Results
- Cross-lingual Semantics (Michael Ellsworth)
    - Web Annotation Tool (Matos and Torrent, 2016)
    - Parallel Annotation on the TED Talk
    - Cross-lingual Framing Differences
    - Metaphor and Metonymy in FrameNet
    - Frame-based Knowledge Representation
    - Mental Spaces: Negation and Conditionals
- ASRL and Multilingual FrameNet Applications (Swabha Swayamdipta)
    - Automatic Frame Semantic Role Labeling (English)
    - Automatic Frame Semantic Role Labeling (Other Languages)
    - Potential Multilingual Applications

3   Instructors

Collin F. Baker

Collin F. Baker (International Computer Science Institute, collinb@icsi.berkeley.edu, `https://www.icsi.berkeley.edu/icsi/people/collinb`) has been affiliated with the FrameNet Project since it began, Project Manager, 2000-present), working closely with the late Charles J. Fillmore and many collaborators. He was also Project Manger of the MetaNet Project at ICSI (2012-2015, `http://metanet.icsi.berkeley.edu`), an IARPA-funded effort to recognize metaphoric language automatically in several languages. His research interests include building and aligning FrameNets across languages (Lönneker-Rodman and Baker (2009), Gilardi and Baker (2017)), aligning FrameNet to other lexical resources (Fellbaum and Baker (2013), (2008), Ferrández et al. (2010)) and linking to ontologies and reasoning (Scheffczyk et al., 2010).

Michael Ellsworth

Michael Ellsworth (Addeco/International Computer Science Institute, (infinity@icsi.berkeley.edu, `https://berkeley.academia.edu/MichaelEllsworth`) has been involved in lexical semantic research for nearly 20 years, and has been a key member of the FrameNet team, involved in frame definition, annotation, annotator training, and data-integrity checking; he has been in charge of the ontology-like hierarchy that organizes the frames since 2002. Publication topics include the differences between FrameNet and other annotation projects (Ellsworth et al., 2004), the FrameNet hierarchy

and ontologies (Dolbey et al., 2006), the principles behind FrameNet annotation, paraphrasing using FrameNet (Ellsworth and Janin, 2007), and various English grammatical constructions (Lee-Goldman et al., 2009). He is currently writing a dissertation on how the domain of emotion is encoded in English words and grammatical constructions.

Miriam R. L. Petruck

Miriam R. L. Petruck (International Computer Science Institute, miriamp@icsi.berkeley.edu, `https://www.icsi.berkeley.edu/icsi/people/miriamp`) received her Ph.D. in Linguistics from the University of California, Berkeley, CA, under the direction of the late Charles J. Fillmore. A key member of the FrameNet development team almost since the project's founding in 1997, her research interests include semantics (Petruck (2009),(1996)), lexical semantics (Petruck and Ellsworth, 2016), knowledge base development, grammar and lexis, semantics, Frame Semantics and Construction Grammar, particularly as these linguistic theories support advances in NLU and NLP. She is a frequent invited speaker, lecturing internationally about Frame Semantics, Construction Grammar, and FrameNet. Petruck is currently working on a manuscript about FrameNet and NLP.

Swabha Swayamdipta

Swabha Swayamdipta (swabha@cs.cmu.edu, `http://www.cs.cmu.edu/~sswayamd`) is a PhD student at the Language Technologies Institute at Carnegie Mellon University (currently a visiting student at U Washington). She works with Noah Smith and Chris Dyer on developing efficient algorithms for broad-coverage semantic parsing, with a focus on exploiting the relationship between syntax and semantics (Swayamdipta et al., 2017). She has a Masters degree from Columbia University, and was a research intern at Google New York. Her research interests also include applications of broad-coverage semantics for tasks such as entailment and coreference.

References

Andrew Dolbey, Michael Ellsworth, and Jan Scheffczyk. 2006. BioFrameNet: A Domain-Specific FrameNet Extension with Links to Biomedical Ontologies. In *Proceedings of KR-MED*.

Michael Ellsworth and Adam Janin. 2007. Mutaphrase: Paraphrasing with FrameNet. In *Proceedings of the Workshop on Textual Entailment and Paraphrasing*, Prague, June. Association for Computational Linguistics.

Michael Ellsworth, Katrin Erk, Paul Kingsbury, and Sebastian Padó. 2004. PropBank, SALSA, and FrameNet: How Design Determines Product. In *Proceedings of the LREC 2004 Workshop on Building Lexical Resources from Semantically Annotated Corpora*, Lisbon, Portugal.

Christiane Fellbaum and Collin F. Baker. 2008. Can WordNet and FrameNet be Made "Interoperable"? In *Proceedings of The First International Conference on Global Interoperability for Language Resources*, pages 67–74.

Christiane Fellbaum and Collin Baker. 2013. Comparing and harmonizing different verb classifications in light of a semantic annotation task. *Linguistics*, 51(4):707–728.

Óscar Ferrández, Michael Ellsworth, Rafael Muñoz, and Collin F. Baker. 2010. Aligning FrameNet and WordNet based on Semantic Neighborhoods. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odjik, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, pages 310–314, Valletta, Malta, may. European Language Resources Association (ELRA).

Charles J. Fillmore. 1977. The need for a frame semantics in linguistics. In Hans Karlgren, editor, *Statistical Methods in Linguistics*. Scriptor.

Charles J. Fillmore. 1985. Frames and the semantics of understanding. *Quaderni di Semantica*, 6(2):222–254.

Luca Gilardi and Collin F. Baker. 2017. Learning to align across languages: Toward multilingual framenet. In Tiago Timponi Torrent, Lars Borin, and Collin F. Baker, editors, *Proceedings of the International FrameNet Workshop 2018: Multilingual FrameNets and Constructicons*. LREC.

Karl Moritz Hermann and Phil Blunsom. 2014. Multilingual models for compositional distributed semantics. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 58–68. Association for Computational Linguistics.

Russell Lee-Goldman, Russell Rhodes, and Michael Ellsworth. 2009. Determination and modification: interaction and interpretation. San Francisco, January. Linguistic Society of America. Presented at annual meeting.

Birte Lönneker-Rodman and Collin F. Baker. 2009. The FrameNet Model and its Applications. *Natural Language Engineering*, 15(03):415–453.

Ely Matos and Tiago Torrent. 2016. A flexible tool for an enriched FrameNet. In *9th International Conference on Construction Grammar (ICCG9)*. Federal University of Juiz de Fora.

Miriam R L Petruck and Michael Ellsworth. 2016. Representing support verbs in framenet. In *Proceedings of the 12th Workshop on Multiword Expressions*, pages 72–77. Association for Computational Linguistics.

Miriam R. L. Petruck. 1996. Frame Semantics. In Jef Verschueren, Jan-Ola Östman, Jan Blommaert, and Chris Bulcaen, editors, *Handbook of Pragmatics*. John Benjamins.

Miriam Petruck. 2009. Typological Considerations in constructing a Hebrew FrameNet. In Hans C. Boas, editor, *Multilingual FrameNets in Computational Lexicography*, pages 183–205. Mouton.

Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, Collin F. Baker, and Jan Scheffczyk. 2016. *FrameNet II: Extended Theory and Practice*. International Computer Science Institute, Berkeley, California. Revised Nov. 1, 2016. Distributed with the FrameNet data.

Jan Scheffczyk, Collin F. Baker, and Srini Narayanan. 2010. Reasoning over Natural Language Text by Means of FrameNet and Ontologies. In Chu-Ren Huang, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci, Alessandro Oltramari, and Laurent Prévot, editors, *Ontology and the Lexicon: A Natural Language Processing Perspective*, Studies in Natural Language Processing, chapter 4, pages 53–71. Cambridge University Press, Cambridge, UK. Expanded version of paper at OntoLex, 2006. (ISBN-13: 9780521886598).

Swabha Swayamdipta, Sam Thomson, Chris Dyer, and Noah A. Smith. 2017. Frame-semantic parsing with softmax-margin segmental rnns and a syntactic scaffold.