SemLink+: FrameNet, VerbNet and Event Ontologies

Martha Palmer, Claire Bonial
Department of Linguistics
University of Colorado
Boulder, CO

mpalmer/Claire.Bonial@colorado.edu

Diana McCarthy
Department of Theoretical and
Applied Linguistics (DTAL)
University of Cambridge

diana@dianamccarthy.co.uk

Abstract

This paper reviews the significant contributions FrameNet has made to our understanding of lexical resources, semantic roles and event relations.

1 Introduction

One of the great challenges of Natural Language Processing (NLP) is the multitude of choices that language gives us for expressing the same thing in different ways. This is obviously true when taking other languages into consideration - the same thought can be expressed in English, French, Chinese or Russian, with widely varying results. However, it is also true when considering a single language such as English. Light verb constructions, nominalizations, idioms, slang, paraphrases, and synonyms all give us myriads of alternatives for "coining a phrase." This causes immense difficulty for NLP systems. No one has made greater contributions to advancing the state of the art of lexical semantics, and its applications to NLP, than Chuck Fillmore. In this paper we focus on the central role that FrameNet has played in our development of SemLink+ and in our current explorations into event ontologies that can play a practical role in accurate automatic event extraction.

2 Detecting events

An elusive goal of current NLP systems is the accurate detection of events – recognizing the meaningful relations among the topics, people, places and events buried within text. These relations can be very complex, and are not always explicit, requiring subtle semantic interpretation of the data. For instance, NLP systems must be able to automatically recognize that *Stock prices sank* and *The stock market is falling* can be describing the same event. Such an interpretation relies upon a recognition of the similarity between *sinking* and *falling*, as well as noting the connection between stock *prices* and the stock *market*, and, finally, acknowledgment that they are playing the same role. A key element in event extraction is the identification of the participants of an event, such as the initiator of an action and any parties affected by it. Basically *who* did *what* to *whom, when, where, why* and *how?* Many systems today rely on semantic role labeling to help identify participants, and lexical resources that provide an inventory of possible predicate argument structures for individual lexical items are crucial to the success of semantic role labeling (Palmer,et al., 2010).

3 SemLink+ and Semantic Roles

SemLink (Palmer, 2009) is an ongoing effort to map complementary lexical resources: PropBank (Palmer et al., 2005), VerbNet (Kipper et al., 2008), FrameNet (Fillmore et al., 2004), and the recently added OntoNotes (ON) sense groupings (Weischedel, et al., 2011). They all associate semantic information with the propositions in a sentence. Each was created independently with somewhat differing goals, and they vary in the level and nature of semantic detail represented. FrameNet is the

Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore (1929–2014), pages 13–17,
Baltimore, Maryland USA, June 27, 2014.

most fine-grained with the richest semantics, VerbNet focuses on syntactically-based generalizations that carry semantic implications, and the relatively coarse-grained PropBank has been shown to provide the most effective training data for supervised Machine Learning techniques. Nonetheless, they can be seen as complementary rather than conflicting, and together comprise a whole that is greater than the sum of its parts. SemLink serves as a platform to unify these resources. The recent addition of ON sense groupings, which can be thought of as a more coarse-grained view of WordNet (Fellbaum, 1998), provides even broader coverage for verbs, and a level of representation that is appropriate for linking between VerbNet class members and FrameNet lexical units, as described below.

SemLink unifies these lexical resources at several different levels. First by providing type-to-type mappings between the lexical units for each framework. For PropBank these are the very coarse-grained rolesets, for VerbNet they are verbs that are members of VerbNet classes, and for FrameNet they are the lexical units associated with each Frame. The same lemma can have multiple PropBank rolesets and can be in several VerbNet classes and FrameNet frames, but always with different meanings. In general, the mappings from PropBank to VerbNet or FrameNet tend to be 1-many, while the mappings between VerbNet and FrameNet are more likely to be 1-1. For example, the verb *hear* has just one coarse-grained sense in PropBank, with the following roleset:

Arg0: hearer
Arg1: utterance, sound
Arg2: speaker, source of sound

This roleset maps to both the *Discover* and *See* classes of VerbNet, and the *Hear* and *Perception_experience* frames of FrameNet.

Then, for each lexical unit, SemLink also supplies a mapping between the semantic roles of PropBank and VerbNet, as well as the roles of VerbNet and FrameNet. PropBank uses very generic labels such as Arg0 and Arg1, which correspond to Dowty's Prototypical Agent and Patient, respectively (Dowty, 1991). PropBank has up to six numbered arguments for core verb specific roles and for adjuncts it has several generally applicable ArgM*odifiers* that have function tag labels such as: M*a*N*ne*R, T*e*MP*oral*, LOC*ation*, DIR*ection*, GO*a*L, etc. VerbNet uses more traditional linguistic thematic role labels, with about 30 in total, and assumes adjuncts (ArgM's) will be supplied by PropBank based semantic role labelers. FrameNet is even more fine-grained and has frame-specific core and peripheral roles called Frame Elements for each frame, amounting to over 2000 individual Frame Element types. For example, *He talked about politics* would receive the following semantic role labels from each framework.[1]

PropBank (talk.01)
He_{Arg0} $\text{talked}_{\text{RELATION}}$ about $\text{politics}_{\text{Arg1}}$

VerbNet (Talk-37.5):
He_{AGENT} $\text{talked}_{\text{RELATION}}$ about $\text{politics}_{\text{TOPIC}}$

FrameNet (Statement frame):
$\text{He}_{\text{SPEAKER}}$ $\text{talked}_{\text{RELATION}}$ about $\text{politics}_{\text{TOPIC}}$

Thanks to Chuck Fillmore's careful guidance, the rich, meticulously crafted Frames in FrameNet, with their detailed descriptions of all possible arguments and their relations to each other, offer the potential of providing a foundation for inferencing about events and their consequences. In addition FrameNet has from the beginning been inclusive in its addition of nominal and adjectival forms to the Frames, which greatly increases our coverage of all predicating elements (Bonial, et al., 2014). There is also a comprehensive FrameNet Constructicon that painstakingly lists many phrasal constructions, such as "the Xer, the Yer" that cannot be found anywhere else (Fillmore, et al., 2012). Many of these frames, including the constructions, apply equally well to other languages, as evidenced by the various efforts to develop FrameNets in other languages[2] promising a likely benefit to multilingual information

[1] Arg0 maps to Agent maps to Speaker. Arg1 maps to Topic maps to Topic.

[2] See FrameNet projects in other languages listed at https://framenet.icsi.berkeley.edu/fndrupal/framenets_in_other_languages

processing as well. Given the close theoretical ties between PropBank, VerbNet and FrameNet, it should be possible to bootstrap from the successful PropBank-based automatic semantic role labelers to equally accurate FrameNet and VerbNet annotators, and to improve overall semantic role labeling performance (Bauer & Rambow, 2011; Dipanjan, et al., 2010; Giuglea & Moschitti, 2006; Merlo & der Plas, 2009; Yi, et al., 2007). That is one of the primary goals of SemLink.

The first release of SemLink (1.1) contained mappings between these three lexical resources as well as a set of PropBank instances from the Wall Street Journal data with mappings to VerbNet classes and thematic roles (Palmer, 2009). Our most recent release, SemLink 1.2,[3] now includes mappings to FrameNet frames and Frame Elements wherever they are available (FN version 1.5), as well as ON sense groupings (Bonial, et al., 2013). The mapping files between PropBank and VerbNet (version 3.2), and FrameNet have also been checked for consistency and updated to more accurately reflect the current relations between these resources.

This annotated corpus can now be used to train and evaluate VerbNet Class and FrameNet Frame classifiers, to explore clusters of Frame Elements that map to the same VerbNet and PropBank semantic roles, and to evaluate approaches to semantic role labeling that use the type-to-type mappings to bootstrap VerbNet and FrameNet role labels from automatic PropBank semantic role labels.

4 Events, Event Types and Subevents

Accurate and informative semantic role labels are an essential component of event extraction, but, although necessary, they are not sufficient. Automatic event detection also requires the ability to distinguish between events which are truly separate, such as *Yesterday, John was throwing a ball to Mary and Bill was flying a kite*, as opposed to related events such as *John was washing the dishes and Mary was drying them.* The second pair could be seen as temporally related subevents of an overall *doing the dishes* or *cleaning up the kitchen* event. It can sometimes be quite challenging to determine the relationship between two events. For instance, earthquakes are quite often associated with the collapse of buildings, as in the following example, *The quake destroyed parts of Sausalito. All tall buildings were demolished.*

Many readers might agree that the *earthquakes* CAUSED the *demolishment of the buildings.* However, are the *building* collapses also SUBEVENTs of the *earthquakes*? Sometimes they happen a few days later, or immediately, simultaneously with the earthquake. Are they both subevents? In general, for accurate event detection, it would be very useful to know which events must precede, must follow, or cannot be simultaneous with, which other events. As discussed in the 2013 NAACL Events workshop and this year's ACL Events workshop, clear, consistent annotation of events and their coreference and causal and temporal relations is a much desired but very challenging goal (Ikuta & Palmer, 2014). Any assistance that can be provided by lexical resources is welcome.

Another very important contribution that FrameNet has made is in the realm of defining these kinds of relations, and others, between frames. Parent-Child Frame to Frame relations can include Inheritance, Subframe, Perspective On, Using, Causative Of, Inchoative of, and there is also a Precedes temporal ordering relation.

The DEFT working group in Richer Event Descriptions has recently been exploring expanding the ACE and ERE event types, and how they can be mapped onto a broader ontological context. Exploring the FrameNet relations that the relevant lexical items participate in has been most informative. We first examined the simple LDC ERE classification of Conflict events, which has *demonstrations* and *attacks* as siblings (ERE guidelines). We find FrameNet's classification of *attacks* as Hostile-Encounters quite useful, and have no argument with it having an Inheritance relation with Intentionally_act, and a Using relation with Taking_sides. *Demonstrations*, on the other hand, come under the Protest Frame, which has a Using relation with Taking_sides. The FrameNet

[3] available for download here: http://verbs.colorado.edu/semlink/

organization of *demonstrations* and *attacks*, although perfectly justifiable, doesn't map neatly onto the LDC organization since, although they are close, they are not siblings. However, by also considering SUMO (Niles & Pease, 2001), the Predicate Matrix (de Lacalle , et al., 2014), WordNet and VerbNet, we were able to develop the upper level partial Event Ontology given in Figure 1, which comfortably incorporates the ERE and FrameNet relations within a broader framework, preserving the key aspects of each.

We are now discussing the ERE Life events, *birth, death, injury, marriage, divorce,* etc., and FrameNet is again proving to be inspirational. SemLink+ will encompass our growing Event Ontology, as well as the mappings between the resources and the multiple layers of annotation on the same data.

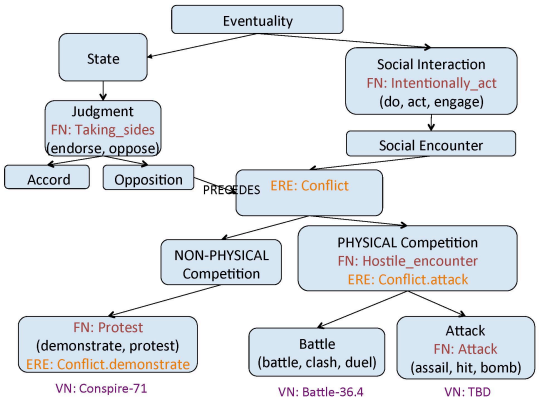

Figure 1 – SemLink+ Event Ontology, partial

5 Conclusion

Since computers do not interact with and experience the world the same way humans do, how could they ever interpret language describing the world the same way humans do? That NLP has made as much progress as it has is truly phenomenal, and there is much more still that can be done. Rich, detailed, lexical resources like FrameNet are major stepping stones that will enable continued improvements in the automatic representation of sentences in context. FrameNet, and WordNet, PropBank, VerbNet and SemLink+, provide priceless, invaluable information about myriads of different types of events and the creative ways in which they can be expressed, as well as rich details about all of their possible participants. If we can harness the power of distributional semantics to help us dynamically extend and enrich what has already been manually created, we may find our computers to be much smarter than we ever imagined them to be.

Acknowledgments

This work has benefited immensely from comments and suggestions during the discussions of the RED working group on Event Ontologies, especially from Teruko Mitamura, Annie Zaenen, Ann Bies, and German Rigau. We also gratefully acknowledge the support of the National Science Foundation Grant NSF-IIS-1116782, A Bayesian Approach to Dynamic Lexical Resources for Flexible Language Processing, DARPA FA-8750-13-2-0045, subaward 560215 (via LDC) DEFT: Deep Exploration and Filtering of Text, DARPA Machine Reading (via BBN), and NIH: 1 R01 LM010090-01A1, THYME, (via Harvard). The content is solely the responsibility of the authors and does not necessarily represent the official views of DARPA, NSF or NIH.

References

Daniel Bauer & Owen Rambow, 2011, Increasing Coverage of Syntactic Subcategorization Patterns in FrameNet Using VerbNet, In the *Proceedings of the IEEE Fifth International Conference on Semantic Computing*.

Claire Bonial, Julia Bonn, Kathryn Conger, Jena Hwang and Martha Palmer, 2014. PropBank: Semantics of New Predicate Types. The *9th edition of the Language Resources and Evaluation Conference.* Reykjavik, Iceland.

Claire Bonial, Kevin, Stowe, and Martha Palmer, 2013. Renewing and Revising SemLink. *The GenLex Workshop on Linked Data in Linguistics,* held with GenLex-13.

Dipanjan Das, Nathan Schneider, Desai Chen, and Noah A. Smith, 2010. Probabilistic Frame-Semantic Parsing. In *Proceedings of the NAACL 2010*.

David Dowty, 1991. Thematic Proto-Roles and Argument Selection. *Language*, 67:547-619

Christiane Fellbaum, 1998. *WordNet: An Electronic Lexical Data-base*. Language, Speech and Communications. MIT Press

Charles. J. Fillmore; Collin F. Baker, and H. Sato, 2004. FrameNet as a ``Net". In *Proceedings of LREC* 2004*, 4*, pages 1091-1094

Charles J. Fillmore, Russell R. Lee-Goldman, and Russell Rhodes. 2012. "The FrameNet Constructicon" Boas, H.C. and Sag, I.A. (Eds.) *Sign-based Construction Grammar*, CSLI Publications.

Ana-Maria Guiglea and Alessandro Moschitti. 2006. Semantic role labeling via FrameNet, VerbNet and PropBank. In Proceedings of Coling-ACL 2006, pages 929–936.

Rei Ikuta and Martha Palmer (2014) Challenges of Adding Causation to Richer Event Descriptions, *In the Proceedings of 2nd Events Workshop, held with ACL 2014,* Baltimore, MD.

Karin Kipper, Anna Korhonen, Neville Ryant and Martha Palmer. 2008. A Large-Scale Classification of English Verbs. *Language Resources and Evaluation Journal, 42(1):21–40*

Maddalen Lopez de Lacalle, Egoitz Laparra, German Rigau, 2014, Predicate Matrix: extending SemLink throughWordNet mappings, The *9th edition of the Language Resources and Evaluation Conference.* Reykjavik, Iceland.

Ian Niles and Adam Pease, 2001. Towards a Standard Upper Ontology. In *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001),* Chris Welty and Barry Smith, eds, Ogunquit, Maine, October 17-19, 2001.

Paola Merlo and Lonneke van der Plas. 2009. Abstraction and generalization in semantic role labels: PropBank, VerbNet or both?, In the *Proceedings of ACL 2009.*

Martha Palmer, 2009. SemLink: Linking PropBank, VerbNet and FrameNet, *In the Proceedings of the Generative Lexicon Conference, GenLex-09*.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics, 31(1):71–106*

Martha Palmer, Daniel Gildea and Nianwen Xue. Semantic Role Labeling. 2010. Synthesis Lectures on Human Language Technology Series, ed. Graeme Hirst, Morgan and Claypoole.

Ralph Weischedel, Eduard Hovy, Mitchell Marcus, Martha Palmer, Robert Belvin, Sameer Pradan, Lance Ramshaw and Nianwen Xue. OntoNotes: A Large Training Corpus for Enhanced Processing, included in Part 1 : Data Acquisition and Linguistic Resources of the *Handbook of Natural Language Processing and Machine Translation: Global Automatic Language Exploitation* Editors: Joseph Olive, Caitlin Christianson, John McCary, Springer Verglag, pp 54-63, 201