

From TreeBank to PropBank

Paul Kingsbury, Martha Palmer

University of Pennsylvania
kingsbur,mpalmer@unagi.cis.upenn.edu

Abstract

This paper describes our approach to the development of a Proposition Bank, which involves the addition of semantic information to the Penn English Treebank. Our primary goal is the labeling of syntactic nodes with specific argument labels that preserve the similarity of roles such as *the window* in *John broke the window* and *the window broke*. After motivating the need for explicit predicate argument structure labels, we briefly discuss the theoretical considerations of predicate argument structure and the need to maintain consistency across syntactic alternations. The issues of consistency of argument structure across both polysemous and synonymous verbs are also discussed and we present our actual guidelines for these types of phenomena, along with numerous examples of tagged sentences and verb frames. Metaframes are introduced as a technique for handling similar frames among near-synonymous verbs. We conclude with a summary of the current status of annotation process.

1. Introduction

Recent years have seen major breakthroughs in natural language processing technology based on the development of powerful new techniques that combine statistical methods and linguistic representations. Yet the goals of accurate information extraction, focused information retrieval and fluent machine translation still remain tantalizingly out of reach. A critical element that is still lacking in current natural language processors is accurate predicate–argument structure. This necessity was clearly demonstrated by a recent evaluation of an English–Korean MT system (Han et al, 2000). Although several factors such as syntactic structure and vocabulary coverage influenced the quality of the translation output, the most important factor was accurate predicate–argument structure. Even with a grammatical parse of the source sentence and complete vocabulary coverage, the translation was often still incomprehensible. The parser may have properly recognized all the constituents that are verb arguments, but if it did not assign precise argument positions to them during the transfer process, or if the argument labels were lost in conversion to the MT system, the wrong constituent was demoted or promoted, or labeled as dropped. All of these seemingly trivial errors produced garbled translations. Simply preserving the proper argument position labels, keeping the same parses and transfer lexicon, resulted in an almost 50% jump in the number of acceptable translations for one of the parsers, from 24% to 33%, and a more than tripling of acceptable translations for the other parser, from 10% to 35%. (Kittredge et al, 2001)

In the same way that the existence of the Penn TreeBanks (Marcus, Santorini, & Marcinkiewicz 1993; Marcus 1994) enabled the development of extremely powerful new syntactic analyzers (Collins 1997, 2000), moving to the stage of accurate predicate argument analysis will require a body of publicly available training data that explicitly annotates predicate argument positions with labels. A consensus on a task–oriented level of semantic representation has been achieved with respect to English, under the auspices of the ACE program. It was agreed that the highest priority, and the most feasible type of semantic annotation, is a predicate–argument structure for verbs, participial modifiers and nominalizations, to be known as Proposition Bank, or

PropBank. This paper describes the PropBank verb predicate–argument structure annotation being done at Penn. Similar projects include Framenet (Baker, Fillmore & Lowe 1998, Gildea 2001) and Prague Tectogramatics (Hajicova, Panevova, & Sgall 2001).

2. Predicate–Argument Structure across Syntactic Frames

The verb of the sentence typically indicates a particular event and the verb’s syntactic arguments are associated with the participants in that event. In the sentence *John broke the window*, the event is a breaking event, with *John* as the instigator and a *broken window* as the result. The associated predicate–argument structure would be **break(John, window)**. Recognition of predicate–argument structures is not straightforward since a natural language will have both several different lexical items which can be used to refer to the same type of event as well as several different syntactic realizations of the same predicate–argument relations. For example, a meeting between two dignitaries can be described using the verbs *meet*, *visit*, *debate*, *consult*, and others¹, each of which are syntactically interchangeable while lending their own individual semantic nuances. Thus, variations such as the following are seen:

- 1) A will [*meet/visit/debate/consult*] (with) B
A and B [*met/visited/debated/consulted*]
*There was a [meeting/visit/debate/consultation]
between A and B*
A had a [meeting/visit/debate/consultation] with B

At the same time, not all syntactic frames of a given verb are interchangeable with those of related verbs:

- 2) Blair [*met/consulted/visited*] with Bush.
The proposal [*met/*consulted/*visited*] with skepticism.

In determining consistent annotations for argument labels of several different syntactic expressions of the same verb, we are relying heavily on recent work in linguistics on word classifications that have a more

¹ These are representative of the *meet* class (36.3) of Levin (1993).

semantic orientation, such as Levin's verb classes (1993), and WordNet (Miller et al, 1990). Levin's classes, and our refinements on them (Dang et al, 1998) provide the key to recognizing the common basis for the myriad ways in which a concept can be expressed syntactically. The verb classes are based on the ability of the verb to occur or not occur in pairs of syntactic frames that are in some sense meaning preserving (diathesis alternations) (Levin 1993). The distribution of syntactic frames in which a verb can appear determines its class membership. The fundamental assumption is that the syntactic frames are a direct reflection of the underlying semantics; the sets of syntactic frames associated with a particular Levin class reflect underlying semantic components that constrain allowable arguments. For example, the following pairs of sentences share many of the same entailments although they are clearly not identical: indefinite object drop, [*We ate fish and chips./ We ate at noon*]; cognates, [*They danced a wild dance./ They danced.*]; and causative/inchoative [*He chilled the soup./ The soup chilled.*]

3. Proposition Bank

The training material for proposition recognition, PropBank, is being annotated in English, based on a consensus developed in 2000 among research groups at BBN, MITRE, New York University, and Penn. Taking as a starting point the Penn Treebank II Wall Street Journal Corpus of a million words (Marcus 1994), we are adding predicate argument structure annotation.² Approximately one-quarter of the TreeBank, comprising largely texts of financial reporting, has been extracted and is serving as our initial focus for training and to provide an earlier delivery of fully-annotated text. This subcorpus should be completed in June of 2002, while the remainder of the corpus will be completed by the summer of 2003. The current project annotates only verbal predicates, setting aside nominalizations, adjectives, and prepositions for a later phase. In a separate paper (Kingsbury, Marcus & Palmer, forthcoming) we discuss the differences between PropBank and similar resources such as Verbnet, Wordnet and Framenet.

In creating annotations for argument structure, a combination of syntactic and semantic factors are used, although syntactic cues are foremost. The general method is the following: for any given predicate, a survey is made of the usages of the predicate and the usages divided into major senses if required. These senses are divided more on syntactic grounds than semantic, thus avoiding the fine-grained and often-arbitrary divisions of, e.g., WordNet. The expected arguments of each sense are then numbered sequentially from Arg0 to Arg5. According to the guidelines established by the ACE community described above, no attempt is made to make argument labels have the same "meaning" from one sense of a verb to another, so for example the "role" played by Arg2 in one sense of a given predicate may be played by Arg3 in another sense. On the other hand, we intend for predicates belonging to the same VerbNet class to share similarly-labeled

arguments, in keeping with the near-synonymy of the predicates.

The examples below demonstrate how argument labels change between senses of one verb ('draw,' examples 4, 5), while a different verb within the same VerbNet class will use the same labels (sense 'pull' examples 4, 7; sense 'art' examples 5, 6).

3) 'draw' sense: **pull**

...the campaign is drawing fire from anti-smoking advocates³...

Arg0:	the campaign
Rel:	drawing
Arg1:	fire
Arg2-from:	anti-smoking advocates
	(role=source)

4) 'draw' sense: **art**

[he was]...drawing diagrams and sketches for his patron

Arg0:	he
Rel:	drawing
Arg1:	diagrams and sketches
Arg2-for:	his patron
	(role = benefactive)

5) 'paint' sense: **art**

...someone using a wide paintbrush could produce a broad line but would have trouble *trace* painting a thin one.

Arg0:	*trace* -> someone
Rel:	painting
Arg1:	a thin one

6) 'pull' sense: **pull**

The EPA will pull a pesticide from the marketplace.

Arg0:	The EPA
Rel:	pull
Arg1:	a pesticide
Arg2-from:	the marketplace
	(role=source)

After we identify the predicate of a clause, we assign appropriate labels to its arguments. These labels are given as normalized argument structure as indicated by (Arg0, Arg1, Arg2, ArgM (for modifiers) plus prepositions). This level of annotation allows us to capture the similarity across syntactic alternations as seen in the following examples:

7) ...the company to offer a 15% stake to the public.

Arg0:	the company
Rel:	offer
Arg1:	a 15% stake
Arg2-to:	the public

8) ...Sotheby's...offered the Dorrance heirs a money-back guarantee

Arg0:	Sotheby's
Rel:	offered
Arg2:	the Dorrance heirs
Arg1:	a money-back guarantee

² BBN has already completed pronoun co-reference annotation on the same data

³ This and all subsequent examples taken from the corpus.

9) ...an amendment offered by Rep. Peter DeFazio...

Arg1: an amendment
Rel: offered
Arg0-by: Rep. Peter DeFazio

10) ...Subcontractors will be offered a settlement...

Arg2: Subcontractors
Rel: offered
Arg1: a settlement

Whenever possible, when transitivity alternations do not occur, we use the same predicate argument structure for all instances of a verb. With *carry*, there are two arguments, *Arg0*, *Arg1* whether a mother is carrying a baby, a bond is carrying a yield, crystals are carrying currents, or viruses are carrying genes. However, occasionally verbs are not used so consistently, as in the case of *leave*. The DEPART sense involves two arguments, *Arg0*, *Arg1* as in *John left the airport* or *John left his wife*, but the GIVE sense requires a third argument, *Arg2*, as in *That would leave Mrs. Thatcher little room for maneuver*, which is characteristic of all verbs of the GIVE class.

We also make use of the existing "functional tags" from the TreeBank, marking nonrequired elements which nevertheless play a role in the event of the verb. These tags include temporals (TMP), locatives (LOC), directionals (DIR), and adverbials of manner (MNR) and purpose (PRP). We further extend this set with tags marking discourse particles and clauses (DIS), causal adverbials (CAU). Modal verbs (MOD) and negation particles (NEG) are also marked in this manner. While these tags normally appear only on nonrequired elements or adjuncts, some tags appear most commonly on numbered arguments, including a marker of "secondary predication" (PRD), indicating that the argument in question acts as a predicate upon some other argument of the same sentence. We include as arguments whatever zero pronouns could be automatically annotated with high accuracy for English. We also make explicit information about tense, modality and negation. All PropBank annotation is done as stand-off annotation pointing to constituents in the original Penn Treebank.

4. Frames Files

In order to ensure consistent annotation, we provide our annotators with detailed and comprehensive examples of all of a verb's syntactic realizations and the corresponding argument labels. These files are part of the PropBank distribution and also available at http://www.cis.upenn.edu/~cotton/cgi-bin/pblex_fmt.cgi. Starting with the most frequent verbs, a series of frames are drawn up to describe the expected arguments, or roles. The general procedure is to examine a number of sentences from the corpus and select the roles which seem to occur most frequently. These roles are then numbered sequentially from Arg0 up to (potentially) Arg5, and each role is given a mnemonic label. These labels tend to be verb-specific, although, following the lead of Framenet, some labels tend to be general to verb classes, while other labels follow the naming conventions of, e.g., theta-role theory. For example, the verb *buy* is expected to have up to five roles:

BUY

Arg0: buyer
Arg1: thing bought
Arg2: seller, bought-from
Arg3: price paid
Arg4: benefactive, bought-for

Rarely, however, will all of these roles occur in a single sentence.⁴ For example:

11) *The company bought a wheel-loader from Dresser.*

Arg0: The company
rel: bought
Arg1: a wheel-loader
Arg2-from: Dresser

12) *TV stations bought "Cosby" reruns for record prices.*

Arg0: TV stations
rel: bought
Arg1: "Cosby" reruns
Arg3-for: record prices.

As much as possible, rolesets are consistent across semantically related verbs. Thus, the *buy* roleset is the same as the *purchase* roleset, and both are similar to the *sell* roleset:

PURCHASE	BUY	SELL
Arg0: buyer	Arg0: buyer	Arg0: seller
Arg1: thing bought	Arg1: thing bought	Arg1: thing sold
Arg2: seller	Arg2: seller	Arg2: buyer
Arg3: price paid	Arg3: price paid	Arg3: price paid
Arg4: benefactive	Arg4: benefactive	Arg4: benefactive

One detail of note is that, in any transaction, the Arg2 "seller" role of buy is equivalent to the Arg0 "seller" role of sell, and vice-versa. An Information Extraction application could use a specific rule shows the mapping between these arguments and their relationship to a "purchase" template. For both Machine Translation and Information Extraction, the buyer and seller need to remain distinct, but for the other applications, such as Information Retrieval, they can be merged into a superset or Metaframe, which could easily be regarded as analogous to the verb roles in the Framenet 'commerce' frameset:⁵

⁴ It is of interest that few verbs exhibit more than three arguments in any syntactic frame, regardless of the number of semantic arguments expected. This suggests a conflict between syntactic and semantic structures, worthy of independent study.

⁵ We are not suggesting here that this Metaframe is identical to the Framenet 'commerce' entry. In particular, Framenet suggests arguments of 'rate' and 'unit' which we do not find syntactically motivated. Nevertheless, the similarities between this metaframe and the Framenet 'commerce' frame act as a nice confirmation of the reality of the framesets.

METAFRAME: Exchange (commodities for cash)⁶

- Arg0: one exchanger
- Arg1: commodity
- Arg2: other exchanger
- Arg3: price paid (cash)
- Arg4: benefactive

Polysemous verbs usually take multiple roletes when the senses require different syntax, and little to no effort is made to make these roletes consistent. For example, *apply* takes three roletes, with little relation to each other:

APPLY

- 1. "ask for"
- Arg0: applier
- Arg1: thing applied for
- Arg2: entity applied to
- example: *Boyer and Cohen ... applying for a patent on their gene-splicing technique...*

- 2. "associate with"
- Arg0: applier
- Arg1: thing applied, associated
- Arg2: applied to
- example: *Gen-Probe ... to apply existing technology to an array of diagnostic products.*

- 3. "smear"⁷
- Arg0: applier
- Arg1: substance
- Arg2: surface
- Arg3: instrument
- example: *"Sterile" maggots could be bought to apply to a wound.*

Because of the desire to make semantically related verbs use the same rolete, roles occasionally can be numbered differently than might be expected. For example, unaccusative verbs start counting from Arg1 rather than Arg0, as do inchoative senses of causative/inchoative verbs:

- | | | |
|--------------|--------|--------------|
| DIE | cf to: | KILL |
| -- | | Arg0: killer |
| Arg1: corpse | | Arg1: corpse |

- | | |
|--|---|
| OPEN (inchoative) | OPEN (causative) |
| -- | Arg0: opener |
| Arg1: thing opening | Arg1: thing opening |
| Ex: <i>The branch of the the Bank for Foreign Economic Affairs opened in July.</i> | Ex: <i>Texas Instruments Inc. opened a plant in South Korea</i> |

Frames are, as of the beginning of April 2002, in place for over 850 verbs, with an average of 30–40 added each week. A combination of the existing frames and other resources such as Verbnets allows these frames to be quickly extended to cover over 1500 verbs, by copying frames from one verb to other members of the same Verbnets class. An early trial of this method took the frames from *destroy* and copied them onto all the other verbs of the same Levin class (class 44), including annihilate, demolish, exterminate, ruin, waste, wreck, and others. Annotation of these verbs confirmed that the copied frames were suitable in almost all cases. The exception was *waste*, which required an additional rolete to account for cases such as *waste (money) on*. Nevertheless, it is encouraging that of the 13 automatically-generated verb frames, only one needed manual correction. *Destroy* is a fairly simple frame, of course, and it is expected that more complicated or polysemous verbs will cause a degradation in the automatic generation of frames.

A similarly interesting case is that of negated and repeated verbs. Repeated verbs such as *re-enter* or *refile* indicate that the action (entering or filing, in these cases) is done again. These frames are very simple to generate, since in all cases seen thus far the repeated verb takes exactly the same framing as the basic verb. This class of verbs is fairly robust within the Treebank. A more complicated situation is that of negated verbs, those which add the prefix *un-* to some other verb. While it might seem that these would also be straightforward adaptations of existing frames, such as *untie* <- *tie*, cases such as *unload* <- *load* are more complicated. While there is certainly a sense of *unload* which is the logical opposite of *load* (load the truck, unload the truck), there is a sense of *unload*, as in *The thrift is unloading its junk-bond portfolio*, which does not have a corresponding sense for *load*. Similarly, *unravel* is semantically the same as *ravel*, not the opposite. These and similar cases make the *un-*verbs poor candidates for automatic generation of frames. Fortunately, they are quite rare within the Treebank.

It is estimated that the whole of the TreeBank contains approximately 3500 unique verbs.

A number of annotators, mostly undergraduates majoring in linguistics, extend the templates in the frames to examples from the corpus. The rate of annotation is between that of POS tagging and syntactic parsing, running around 50 sentences per annotator-hour. The learning curve for the annotation task is very steep, with most annotators requiring about three days to achieve a degree of confidence and complete independence from the trainer. Difficulties after this period tend to stem from highly marked and very strange syntactic constructions, or from inconsistencies within the syntactic parsing given by Treebank.⁸ These

⁶ This superset is almost identical to the frameset for "trade."

⁷ This could be regarded as a specialization of the "associate with" rolete, but since it can take an instrument, while "associate with" cannot, it is fruitful to regard it as completely separate. Penn's related project Verbnets works on establishing more exactly the relationships between verb senses.

⁸ This is not disparage the accuracy of the Treebank, or to suggest that the parse is not a crucial starting point for our task. Rather, the task itself, by cutting through the corpus at a different angle, highlights the inherent inconsistencies in any treatment of natural language.

"Treebank errors" are being tagged in the hopes of correcting them at some future point.

Inter-annotator agreement is generally high, varying from verb to verb and annotator to annotator, but usually running between 80 and 100 percent. Agreement is calculated by tagged constituent, so a sentence with three arguments (plus the verb itself), for example, counts as four data points. Errors tend to be systematic rather than random, indicating temporary misunderstandings about tagging practices rather than real disagreement over the proper tagging of a given sentence. As such, they are easily caught and corrected.

5. Conclusion

We have presented our basic approach to creating Proposition Bank, which involves adding a layer of semantic annotation to the Penn Treebank. Without attempting to confirm or disconfirm any particular semantic theory, our goal is to provide consistent argument labeling that will facilitate the automatic extraction of relational data. In order to ensure reliable human annotation, we provide explicit guidelines for labeling all of the syntactic and semantic frames of each particular verb. Our rate of progress and our inter-annotator agreement figures demonstrate the feasibility of the task.

6. Acknowledgments

We gratefully acknowledge guidance from Mitch Marcus and Christiane Fellbaum, programming support from Scott Cotton, and tireless labor from our many annotators, especially Betsy Klipple, Olga Babko-Malayo, and Kate Forbes. This work has been funded by NSF IIS-9800658, Verbnet, and a Department of Defense Grant, MDA904-00C-2136.

7. References

- Baker, C.F., Fillmore, C.J., & Lowe, J.B. (1998). The Berkeley Framenet Project. In Proceedings of COLING/ACL (pp 86-90). Montreal: Association for Computational Linguistics.
- Charniak, E. (1995). Parsing with Context-Free Grammars and Word Statistics. In Technical Report: CS-95-28, Brown University.
- Collins, M. (1997). Three generative, lexicalised models for statistical parsing. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics. (pp 16-23) Madrid, Spain: Association for Computational Linguistics.
- Collins, M. (2000). Discriminative reranking for natural language parsing. In International Conference on Machine Learning. (pp 175-182) San Francisco: Morgan Kaufmann.
- Dang, H.T., Kipper, K., Palmer, M., & Rosenzweig, J. (1998) Investigating Regular Sense Extensions based on Intersective Levin Classes. In Proceedings of Coling-ACL98. Montreal: Association for Computational Linguistics.
- Gildea, D. (2001). Statistical Language Understanding Using Frame Semantics. Ph.D. thesis, University of California at Berkeley, Department of Computer and Information Sciences.
- Hajicova, E., Panevova, J., & Sgall, P. (2001) Tectogramatics in Corpus Tagging. In Perspectives on Semantics, Pragmatics, and Discourse: A Festschrift for Ferenc Keifer, I. Kenesei and R.M. Harnish eds. (pp 294-299) Amsterdam/Philadelphia: John Benjamins Publishing Co.
- Han, C., Lavoie B., Palmer, M., Rambow, O., Kittredge, R., Korelsky, T., Kim, N. & Kim, M. (2000) Handling Structural Divergences and Recovering Dropped Arguments in a Korean/English Machine Translation System. In Proceedings of the Association for Machine Translation in the Americas 2000. (pp 40-53) Berlin/New York: Springer Verlag.
- Kingsbury, P., Marcus, M. & Palmer, M. (forthcoming). Adding Predicate Argument Structure to the Penn TreeBank. In Proceedings of the Human Language Technology Conference, San Diego, CA, March 2002.
- Kittredge, R., Korelsky, T., Lavoie, B., Palmer, M., Han, C., Park, C., & Bies, A. (2001) SBIR-II: Korean-English Machine Translation of Battlefield Messages. Final Report to the Army Research Lab.
- Levin, B. (1993) English Verb Classes and Alternations A Preliminary Investigation. Chicago: University of Chicago Press.
- Marcus, M. (1994). The Penn treebank: A revised corpus design for extracting predicate-argument structure. In Proceedings of the ARPA Human Language Technology Workshop, Princeton, NJ.
- Marcus, M., Santorini, B., & Marcinkiewicz, M.A. (1993). Building a large annotated corpus of English : the Penn Treebank. Computational linguistics 19(2), 313-330
- Miller, G., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1993) Five papers on wordnet. International Journal of Lexicography 3(4), 235-312.