Assigning XTAG Trees to VerbNet

Neville Ryant, Karin Kipper
University of Pennsylvania
200 South 33rd Street
Philadelphia, PA 19104 USA
nryant@unagi.cis.upenn.edu
kipper@linc.cis.upenn.edu

Abstract

This paper presents the mappings between the syntactic information of our broad-coverage domain-independent verb lexicon, VerbNet, to Xtag trees. This mapping between complementary resources allowed us to increase the syntactic coverage of our verb lexicon by capturing transformations of the basic syntactic description of the verbs present in VerbNet. In addition, having these two resources mapped allows the semantic predicates present in our lexicon to be used to disambiguate Xtag verb senses.

1 Introduction

The limited availability of large-scale lexical resources has restricted natural language applications to specific domains. We propose to fill this gap by creating VerbNet (Kipper et al., 2000a; Dang et al., 2000), a freely available broad-coverage verb lexicon. VerbNet includes mappings to other known resources so that they can be used as extensions of each other.

VerbNet is a domain-independent verb lexicon with explicit syntactic and semantic information for over 4,000 English verbs. The verbs are organized in classes according to Levin's classification (Levin, 1993). In order to retain common syntactic and semantic properties for all members of a class, our verb classes are hierarchically organized, with 74 new subclasses added to the original classes. The syntactic frames represent the surface structure of constructions such as transitives, intransitives, prepositional phrases, resultatives, and other alternations listed in Levin.

The verbs in our lexicon have been mapped to WordNet (Miller, 1985; Fellbaum, 1998) and more recently to FrameNet (Baker et al., 1998). The syntactic coverage of VerbNet has been tested against the frames found in PropBank (Kingsbury and Palmer, 2002) through a systematic mapping between the two resources. The syntactic frames in our verb lexicon account for over 78% exact matches to the frames found in PropBank (Kipper et al., 2004).

A natural extension of VerbNet's syntactic frames is to incorporate the possible transformations of each frame. The Xtag grammar (XTAG Research Group, 2001) presents a large existing grammar for English verbs that accounts for just that richness of constructions. Mapping our syntactic frames to the Xtag trees greatly increases the robustness of our resource by capturing such transformations.

2 Levin Classes

Levin verb classes (Levin, 1993) are based on the ability of a verb to occur in pairs of syntactic alternations which preserve the intended meaning. The fundamental assumption of Levin classes is that the syntactic behavior of verbs is a direct reflection of the underlying semantics. This is a not uncontroversial thesis in its strongest form, but it is indisputable that meaning can have great predictive ability. Hale and Keyser (1987) discuss the predictive ability of lexical semantic knowledge using the archaic whaling term *gally*, which might be interpreted to mean *see* or possibly *frighten*. Depending on the assumption made about galley's meaning, speakers can make conflicting judgments about the verb's syntactic behavior. For the speaker interpreting it to mean *see*, the middle construction is disallowed "*Whales gally easily" (paralleling "Whales see easily"), while for the speaker who interprets it as *frighten*, the middle construction is allowed. For an example from Levin, consider the classes of the *break* verbs and the *cut* verbs which are similar in the ability of their members to participate in the transitive and middle constructions. Additionally, *break* verbs may appear in the simple intransitive construction while *cut* verbs may appear in the conative construction. The ex-

TAG+7: Seventh International Workshop on Tree Adjoining Grammar and Related Formalisms.
May 20-22, 2004, Vancouver, BC, CA.
Pages 194-198.

planation given by Levin is that the *cut* verbs describe a series of actions with the goal of separating some entity into pieces. Whether the goal is achieved or not, the action can still be performed, as recognized by "John cut at the loaf." For the break verbs, the verb specifies the manner that a resultant change of state occurs. The action of breaking cannot be attempted if no result is achieved and so these verbs disallow "*John broke at the window".

3 VerbNet

VerbNet is a verb lexicon with syntactic and semantic information for English verbs, referring to Levin verb classes (Levin, 1993) to construct the lexical entries. It exploits the systematic link between syntax and semantics that motivates these classes, and thus provides a clear and regular association between syntactic and semantic properties of verbs and verb classes (Kipper et al., 2000a; Dang et al., 2000). Each node in the hierarchy is characterized extensionally by its set of verbs, and intensionally by syntactic and semantic information about the class and a list of typical verb arguments. The argument list of each entry consists of thematic labels and possible selectional restrictions on the arguments expressed using binary predicates. The syntactic information in each verb's entry maps the list of thematic arguments to the deep-syntactic arguments of that verb (normalized for voice alternations, and transformations). The semantic predicates list the participants during various stages of the event described by the syntactic frame.

The syntactic frames act as a short-hand description for the surface realizations allowed for the members of the class. They describe constructions such as transitive, intransitive, prepositional phrase complement, resultative, and a large set of Levin's alternations. A syntactic frame consists of the verb itself, the thematic roles in their preferred argument positions around the verb, and other lexical items which may be required for a particular construction or alternation. Additional restrictions may be further imposed on the thematic roles (quotation, plural, infinitival, etc.). Examples of syntactic frames are *Agent V Patient* (e.g., John hit the ball), *Agent V at Patient* (e.g., John hit at the window), and *Agent V Patient[+plural] together* (e.g., John hit the sticks together).

The semantic information for the verbs is expressed as a conjunction of semantic predicates, such as *motion, contact, transfer_info*. For the same verb, each different alternation typically has a slightly different set of semantic predicates, although there is usually a substantial overlap within a class. The predicates can take arguments over the verb complements, as well as over implicit existentially quantified event variables.

4 Compositional Semantics for VerbNet

Several attempts have been made to use LTAG derivation trees to compute compositional semantics. Stone and Doran (1997) describe a system for incorporating semantics into TAG trees by a system that simultaneously constructs the semantics and syntax of a sentences using LTAGS. Each lexical item anchors a tree or family of trees and associates with each tree a logical form representing the semantic and pragmatic information for that lexical item and tree. The meaning of a sentence is computed by the conjunction of the meaning of the elementary trees used in the derivation.

Joshi and Vijay-Shanker (1999) and Kallmeyer and Joshi (1999) describe the semantics of the derivation tree as a set of attachments to trees. For each attachment, the semantics are defined as a conjunction of formula in a flat semantic notation. They provide an explicit methodology for composing semantic representations.

Kipper et al (2000) present a method for deriving compositional semantic interpretations from sentences using VerbNet. The mappings discussed here are a step closer to that proposal.

5 Extending VerbNet with XTAG

VerbNet, while providing an explicitly constructed verb lexicon with syntax and semantics, offers limited syntactic coverage since it describes only the declarative frame for each syntactic construction or alternation. The Xtag grammar, on the other hand, is a lexical resource with well-characterized syntactic descriptions for lexical items but makes no distinctions between verb senses and currently has contains no explicit semantics. An obvious way to extend VerbNet's syntactic coverage is to incorporate the coverage of Xtag, accounting for the possible transformations of each declarative frame. Presumably, transformations of VerbNet's syntactic frames are recoverable by mapping onto elementary trees of TAG tree families. Then, for any verb in VerbNet each thematic role can be mapped to an indexed node in the basic syntactic tree and the selectional restrictions on VerbNet thematic roles to features on the nodes. In addition to increasing the coverage of VerbNet, this provides us with a pre-existing parser for computing derived and derivation trees to which our semantic predicates can be added and therefore sense distinctions can be made more explicit.

5.1 Mapping VerbNet frames to XTAG

Each frame in VerbNet is described by 4 components: 1) a brief text description (such as *Transitive, Resultative*), 2) an example sentence, 3) a syntactic frame, 4) a semantic description using a set of semantic predicates. Text descriptions and syntactic frames are very much interrelated, but the text description is independent of the roles assigned to the verb's arguments. These text descriptions consist of both primary and secondary descriptions which were made completely consistent for the whole VerbNet lexicon prior to these mappings. Examples of primary descriptions include *Transitive, Material/Produce Alternation, and Ergative*. Secondary descriptions provide additional information about the semantics and/or syntax. These might specify the types of prepositional phrases that a verb may take or the existence of restrictions on a complement (often secondary descriptions are used to distinguish between different types of sentential complements). Examples (1) and (2) show how the first three components of VerbNet frames are described:

(1) *Material/Product Alternation Intransitive (Material Subject)*
"That acorn will grow into an oak tree."
Material V Prep(into) Product

(2) *Material/Product Alternation Intransitive (Product Subject)*
"An oak tree will grow from that acorn."
Product V Prep(from out of) Material

Because secondary descriptions sometimes refer to variants of a frame that correspond in Xtag to an entirely different tree family than the original frame, it is necessary to consider both these descriptions in order to uniquely identify frames. For instance the Benefactive Alternation in VerbNet has two variants as shown in examples (3) and (4):

(3) *Benefactive Alternation (for variant)*
"Martha carved a piece of wood for the baby"
Agent V Material Prep(for) Beneficiary

(4) *Benefactive Alternation (double object)*
"Martha carved the baby a toy out of a piece of wood"
Agent V Beneficiary Product Prep(from out of) Material

In the current Xtag grammar example (3) corresponds to the tree family of Ditransitives with a PP complement (Tnx0Vnx1pnx2), derived from the simple Transitive tree family with the PP anchored by *for* adjoined into the tree at the VP node, whereas example (4) corresponds to the Ditransitive tree family (Tnx0Vnx2nx1). VerbNet however can only discriminate between the two frames with both the primary and secondary descriptions. Each syntactic frame, then, is assumed to be uniquely specified by its primary and secondary descriptions. Generally, the VerbNet syntactic frame specified by a full description corresponds to the surface syntactic realization of an Xtag elementary tree. Mappings between VerbNet syntactic frames and Xtag tree families were done manually, using the latest frozen release of the XTAG grammar and the latest version of VerbNet. Each VerbNet syntactic frame was mapped to a corresponding Xtag tree family, with the index of the tree family recorded in the VerbNet entry. In theory we should be able to annotate each unique VerbNet syntactic frame with a mapping to an Xtag elementary tree. However, there currently are two impediments to doing this:

1. Many VerbNet syntactic frames specify surface realizations of trees that would not be regarded as initial trees in the Xtag framework (though it is possible to regard them as such by violating certain fundamental assumptions of the grammar). The canonical example is where VerbNet includes as part of a frame a PP that Xtag would analyze as an adjunct.

2. Not all VerbNet syntactic frames correspond to an Xtag elementary tree.

The first issue includes certain verbs appearing in the *Induced Action* alternation, for example, and many of the transitive frames that additionally specify a path PP to indicate the direction of the action. In Example (5) the Xtag grammar analyzes 'over the fence' as an adjunct, this analysis is based on the fact that this PP is optional for the grammaticality of the sentence. Consequently, verbs taking this frame should map to the Transitive tree family, the tree corresponding to the VerbNet frame's overt syntax being derived by adjunction into the elementary tree of the auxiliary tree of 'over the fence.' As an example, consider the two frames in (5) and (6), both of which have PP adjuncts under the Xtag analysis:

(5) *Induced Action (with accompanied motion and path PP)*
"Tom jumped the horse over the fence"
Agent V Theme Prep[+spatial] Location

(6) *Transitive (+ path PP)*
"Jackie accompanied Rose to the store" Agent V Theme Prep[+loc OR +path]

This is similarly an issue with intransitives followed by a PP. Xtag grammar guidelines specify that no verb should appear both in Tnx0v (the tree family for purely intransitive verbs that can be followed by a prepositional phrase but do not require one to be grammatical) and also in Tnx0Vpnx1 (the tree family for intransitive verbs that must be followed by a prepositional phrase to be grammatical). In VerbNet many verbs participate in the *Conative* Alternation, in which the a transitive frame alternate with an intransitive frame in which the NP object is replaced with a PP fronted by 'at.' Examples (7) and (8) show a conative frame in VerbNet and its transitive equivalent respectively:

(7) *Conative*
"Carol cut at the bread"

Agent V at Patient

(8) *Basic Transitive*
"Carol cut the bread"
Agent V Patient

While most of the VerbNet verbs that take the conative do not have intransitive forms that are grammatical when not followed by a PP, many can appear in such frames. For example, the verbs *cut, hack, hew, scrape, scratch, shovel, and dust* constitute a partial listing of the verbs taking the conative and that can also appear as bare intransitives. There remains the question, then, of how the conative alternation should be handled by VerbNet. In instances such as these, as we are interested merely in recovering the various transformationally related forms of frames, we simply ignore the constraints of the Xtag grammar, in the case of (5) and (6) mapping to the elementary tree anchoring tree family Tnx0Vnx1Pnx2, the tree family of verbs taking an NP complement followed by a PP complement headed by a particular preposition. In the case of (7) and (8), mapping to Tnx0Vpnx1, the tree family of intransitives with PP complements.

With regard to the second of these considerations, not all VerbNet syntactic frames correspond to some Xtag elementary tree. A number of VerbNet classes contain syntactic frames that specify multiple adjuncts. As a case in point consider VerbNet class `turn-26.6`, with members *alter, metamorphose, transform, transmute, change, convert, and turn*. Each of these can appear in the two frames presented in (9) and (10):

(9) *Causative/Inchoative Alternation (causative, + Material + Product)*
"The witch turned him from a prince into a frog"
Agent V Patient Prep(from) Material Prep(into) Product

(10) *Causative/Inchoative Alternation (inchoative, + Material + Product)*
"He turned from a prince into a frog"
Patient V Prep(from) Material Prep(into) Product

In the Xtag grammar, the frame presented in (9) corresponds to no elementary tree of any tree family. One might disagree over what elementary tree it is derived from. For instance, (9) can be seen as a transitive sentence with PP adjuncts (and thus belonging to tree family Tnx0Vnx1), as a ditransitive taking a PP complement with another PP adjunct (tree family Tnx0Vnx1pnx2), or as a resultative with a PP anchor and an additional PP adjunct (and thus belonging to tree family TRnx0Vnx1Pnx2). Similarly, the frame presented in (10) can be seen as either an intransitive sentence with optional PP adjuncts (Tnx0V), or as a resultative with ergative verb and PP anchor (TREnx1VPnx2). In the current version of VerbNet we have 18 syntactic frames that fall into this category. Some of these are frames that have been added to VerbNet during attempts at expanding syntactic coverage. Others, such as the *middle construction*, are based on the original alternations proposed by Levin. Currently, these frames are also mapped to the Xtag elementary tree from which they are derived, but it is noted that they are not initial trees. In the future, some of these frames may be removed (in the cases where they are not crucial to characterizing the structure of classes) and for others, it should be specified if they are derived from an initial tree.

5.2 Coverage

As of the latest release of VerbNet, there are 196 unique frames (as distinguished by primary and secondary description). Of these, all but 18 correspond exactly to some Xtag elementary tree (the exceptions are discussed above). For these 168 VerbNet syntactic frames that map exactly to an Xtag elementary tree, only 16 of the 57 Xtag elementary trees were used. A detailed inspection on the 41 Xtag tree families with no corresponding VerbNet frame, revealed that 22 of them deal with small clauses, 8 with idiomatic expressions, and 9 with other various classes. However, some of VerbNet's syntactic frames quite simply are not able to be parsed by the Xtag grammar. The current Xtag analysis for PPs analyzes PP complements with an expanded PP structure rather than as a PP substitution node contrasts these approaches. This is done as expansion of the PP makes the NP node of the PP available to the metarules for creating the trees for extraction so that sentences such as (12) are derivable from (11).

(11) *Jill placed her handbag on the table.*

(12) *What 1 did Jill put her handbag on t1?*

However, Xtag's explicit realization of NPs in complement PPs precludes handling of incidences of exhaustive PP substitution. Thus, Xtag does not handle verbs that take an exhaustive PP such as 'here' (or 'there,' 'somewhere,' etc) as an argument. As such, sentences such as (13) currently cannot be handled, and therefore certain frames in VerbNet (namely, the Transitive (+ here/there)) construction simply have no Xtag mapping.

(13) *I spooned the sauce there.*

6 Conclusion

We presented a detailed account of our mappings between our broad-coverage verb lexicon with explicit semantics, VerbNet, and a syntactically rich lexical resource, the Xtag grammar. By incorporating the transformations of the basic frames from Xtag to our syntactic frames in VerbNet we are able to greatly increase the robustness of our resource by indirectly providing a much larger syntactic coverage. In addition to increasing the coverage of VerbNet, these mappings supply us with a pre-existing parser for computing derived and derivation trees to which our semantic predicates can be associated thus helping the task of verb sense disambiguation.

Acknowledgements

This research was partially supported by NSF Grants IIS-9900297, IIS-0325646, DARPA Tides Grant N66001-00-1-891 and ACE Grant MDA904-00-C-2136.

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the 17th International Conference on Computational Linguistics (COLING/ACL-98)*, pages 86–90, Montreal. ACL.

Hoa Trang Dang, Karin Kipper, and Martha Palmer. 2000. Integrating compositional semantics into a verb lexicon. In *Proceedings of the Eighteenth International Conference on Computational Linguistics (COLING-2000)*, Saarbrücken, Germany, July-August.

Christiane Fellbaum, editor. 1998. *WordNet: An Eletronic Lexical Database*. Language, Speech and Communications. MIT Press, Cambridge, Massachusetts.

Kenneth Hale and Samuel Jay Keyser. 1987. *A view from the Middle*. Lexicon Project Working Papers 10, MIT, Cambridge, MA.

Aravind K. Joshi and K. Vijay-Shanker. 1999. Compositional Semantics with Lexicalized Tree-Adjoining Grammar: How Much Under-Specification Is Necessary? . In *Proceedings of the Third International Workshop on Computational Semantics (IWCS-3)*, pages 131–145, Tilburg, The Netherlands, January.

Laura Kallmeyer and Aravind K. Joshi. 1999. Underspecified Semantics with LTAG . In *Proceedings of Amsterdam Colloquium on Semantics*.

Paul Kingsbury and Martha Palmer. 2002. From treebank to propbank. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, Las Palmas, Canary Islands, Spain.

Karin Kipper, Hoa Trang Dang, and Martha Palmer. 2000a. Class-based construction of a verb lexicon. In *Proceedings of the Seventh National Conference on Artificial Intelligence (AAAI-2000)*, Austin, TX, July-August.

Karin Kipper, Hoa Trang Dang, William Schuler, and Martha Palmer. 2000b. Building a class-based verb lexicon using tags. In *Proceedings of the Fifth International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+5)*, pages 147–154, Paris, France, May.

Karin Kipper, Benjamin Snyder, and Martha Palmer. 2004. Extending a verb-lexicon using a semantically annotated corpus. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC-04)*, Lisbon, Portugal.

Beth Levin. 1993. *English Verb Classes and Alternation, A Preliminary Investigation*. The University of Chicago Press.

George Miller. 1985. Wordnet: A dictionary browser. In *Proceedings of the First International Conference on Information in Data*, Waterloo, Ontario.

Matthew Stone and Christine Doran. 1997. Sentence Planning as Description Using Tree Adjoining Grammar. In *Proceedings of ACL-EACL '97*, Madrid, Spain.

XTAG Research Group. 2001. A lexicalized tree adjoining grammar for english. Technical Report IRCS-01-03, IRCS, University of Pennsylvania.