Representing Verb Alternations in Wordnet

Karen T. Kohl, Robert C. Berwick,
Douglas Jones, and Naoyuki Nomura

MIT Center for Biological and Computational Learning
Cambridge, MA 02139

Abstract

The Wordnet enterprise, as George Miller has so aptly put it, forges "the passage from computing with numbers to computing with words." In this chapter we show how the Wordnet framework can be extended to further the marriage of lexicography—traditional dictionaries— and that extra knowledge

But they have persistent problem. When they look up a word, especially a commonly used word, they often find a dozen or more different meanings. What the dictionary does not make clear are the contexts in which each of these different meanings would be understood. So we know what kind of information is required, but we have not yet learned how to provide it to a computer.

(G. Miller, U.S./Japan Joint Workshop on Electronic Dictionaries and Language Technologies January 23–25, 1993.)

Suggests three basic enhancements to Wordnet; each is covered in this chapter: first, Wordnet should say what *is not* a possible sentence pattern as well as what *is* a possible sentence pattern, and do so in the context of actual sentences, rather than just frame-like patterns. To do this, we have added Second, Worndnet should be multilingual: b alternations (what is and is NOT possible to say); second, multilingual Wordnet cross-linguistic comparisons

In this paper, we describe enhancements to two electronic dictionary resources: (1) We have annotated the 2600 verbs of Part One of Beth Levin's

English Verb Classes and Alternations (EVCA) with WordNet word senses, and (2) we have developed resources for generating sample sentences for verbs in WordNet. In particular, we have built a system that generates 9211 example sentences that a modified version of WordNet reads, where the verbs of EVCA have been folded into our modified WordNet.

1 Introduction

1.1 Project Overview

Our group wants to add the information of the 213 sentence patterns of the 2600 verbs in Beth Levin's *English Verb Classes and Alternations* (EVCA) to WordNet. To improve the sample sentences of verbs in WordNet and to increase the number of sentence patterns, we used many of Levin's verbs and alternation patterns in the Part One of EVCA to build a generation system in Prolog for these real example sentences. In addition to fine-tuning, we give negative examples as seen below.

> BAD: The planet revolves the sun.

These examples are important to know because a similar verb can be in the same pattern, but the sentence may be grammatical as n the below sentence.

> The planet circles the sun.

We generated sample sentences for all of the verbs in Part One of EVCA. EVCA itself gives only one sentence per verb class, so significant effort was required to produce natural sentences for each verb. The Prolog notation for an EVCA verb class can be seen below.

```
evca_dataset(2,
     [coil-3, revolve-2, rotate-1, spin-1, turn-2, twirl-1, twist-5, whirl-1,
     wind-3],
     [pattern(7:ii,'Motion Around an Axis',
          [
           eg(12:a,s,1,
              'Janet broke the cup.',
              [np,v,np]),
           eg(12:b,s,1,
              'The cup broke.',
              [v,np])]),
      pattern(105:ii,'Verbs of Motion Around an Axis',
          [
           eg(106:a,s,1,
              'The spaceship revolves around the earth.',
              [v,np,[p(around,1),np]]),
           eg(106:b,s,0,
              'The spaceship revolves the earth.',
              [v,np,[p(around_0,1),np]])])]).
```

If we had taken the noun phrases from Levin's example sentences and substituted these noun phrases blindly for any verb in the same class, we would not have come up with natural sentences. For example, the example alternation (12) for the Causative/Inchoative Alternation can be seen below.

a. Janet broke the cup.
b. The cup broke.

The verbs *bend*, *crease*, *crinkle*, *crumple*, *fold*, *rumple*, and *wrinkle* also follow the same Causative/ Inchoative Alternation. Blind substitution would have found Janet folding the cup although this is not a possible action or a natural sentence.

1.2 English Verb Classes and Alternations

Part One of Beth Levin's *English Verb Classes and Alternations* (EVCA) explores both the syntax and semantics of more than three thousand verbs.

The verbs are classified according to similar meaning and similar behavior in sentence alternation patterns. Levin suggests that a verb's behavior in sentence alternations depends on its meaning. In EVCA, Levin gives real example sentences with each verb class to demonstrate the alternation. In the first half of EVCA, there are 2600 verbs and 213 sentence patterns after our group's modifications. A single sentence pattern may be grammatical for one class of verbs and ungrammatical for another. Section 1.1.2.1 of EVCA gives the verb class (7) with the alternation (12). Section 1.4.1 gives the same verb class (105) with the alternation (106).

(7) Roll Verbs: bounce, drift, drop, float, glide, move, roll, slide, swing including Motion Around an Axis: coil, revolve, rotate, spin, turn, twirl, twist, whirl, wind

(12) a. Janet broke the cup.
b. The cup broke.

(105) *ROLL VERBS: bounce, drift, drop, float, glide, move, roll, slide swing including MOTION AROUND AN AXIS: coil, revolve, rotate, spin, turn, twirl, twist, whirl, wind

(106) a. The spaceship revolves around the earth.
b. *The spaceship revolves the earth.
(on the interpretation "The spaceship circles the earth.")

1.3 WordNet

WordNet is an on-line dictionary with a thesaurus that was created by a group of psycholinguists at Princeton University.[1] WordNet distinguishes the different senses of words and produces synonyms and sample sentence frames each sense. WordNet was not designed to recognize the syntactic patterns that come from the semantic meaning of the verbs. WordNet does,

[1] See G. A. Miller and C. Fellbaum's *Semantic Networks of English* for a complete description.

however, include at least one generic sentence frame for each sense. These frames distinguish features of the verbs by showing the patterns of sentences that the verbs may take. These sample sentences indicate specific features of verbs, such as argument structure, prepositional phrases and adjuncts, sentential complements, and animacy of the noun arguments.[2] The sentence frames are limited to a simple format[3] that includes the following:

Somebody —-s something PP.
Something is —-ing PP.
Somebody —-s something to somebody.
Somebody —-s somebody.

2 Enhancing WordNet with EVCA Syntactic Classes

One sentence was generated per word sense per alternation pattern. First we parsed by hand the example sentences for each of Levin's verb classes. We then gave the verbs the WordNet sense number. See Appendix B for examples of EVCA datasets, or verb classes, containing the verbs with sense numbers, the example sentences, and the parses of the example sentences. Next, all of the verbs and the noun phrases in their alternations were studied to learn what properties were necessary for each noun phrase in the sentences for a particular verb. Using these properties, specific nouns were created, and these nouns comprised a "toy world" from which the sentences would be generated. One example is using the verb rotate. Not all things can rotate. For something to rotate, it must be solid and axial. Thus the property list for the direct object of *rotate* is *[thing,solid,axial]*, and one instance of a solid, axial thing is a top. These example sentences and property lists would be very useful for learning the language. After comparing the verbs that appear in

[2]See Wendy McKnitt's "A Comparative Analysis of Beth Levin's English Verb Class Alternations and WordNet's Senses for the Verb Classes HIT, TOUCH, BREAK and CUT" for an interesting comparison between WordNet and EVCA.

[3]See Appendix A for a complete listing of the thirty-four sentence frames.

WordNet with those that appear in EVCA, We noticed that while WordNet is very fine-grained in some cases of verbs, some semantic classes are missing from WordNet. For example, the relatively morphologically productive *de-* and *un-* prefixed verbs such as declaw and unzip are largely missing from WordNet, as we shall discuss later.

2.1 Parses of Verb Class Alternations

We hand parsed the sentences in Levin's classes, replacing the lexical items of the sentence with annotated parts of speech labels.[4] These parses, or schema, included such elements as noun phrases, verbs, prepositions, pronouns, and adverbs.[5] We revised the parses several times. Several verbs occurred in more than one dataset, and there needed to be consistency among the parses in the different datasets. The prepositions were not always explicitly stated at first. The parses were revised also to give specific prepositions for each alternation in a class of verbs. Sometimes, however, not all verbs would fit using the same preposition. In these cases, and verb class, or dataset, was split into smaller classes that could all take the same preposition. There ended up a total of 247 datasets.

In trying to find out how many kinds of words are there, we look first at the noun phrases. Each noun phrase has a thematic role, for example: agent, theme, figure, ground, and others. We want to know how to represent these thematic roles and in what parts of a sentence may these thematic roles appear. In the original parses, we gave each noun phrase its explicit thematic role. Then we wanted to derive these patterns of thematic roles from other more basic principles. This would hopefully reduce the number of elements in the schema, changing many explicitly marked noun phrases to simply a noun phrase.

[4]Doug Jones did the initial hand parse, and we modified the parses as the project progressed to suit our needs, for example, to fit the X-bar schema which we discuss in section 2.2

[5]These examples can now be parsed using Brian Ulicny's parser. See Ulicny's "Using English Verb Classes to Parse, Generate and Interpret" in this collection for a description of this parser.

In deriving the thematic roles, we considered the agent, or subject, to be any noun phrase to the left of the verb and the theme, or object, to be any noun phrase not in a prepositional phrase to the right of the verb. Currently noun phrases may act as the subject of the verb phrase, the object of the verb phrase, or the object of a preposition. We are considering adding the subject of a preposition to avoid the question of obligatory adjunction of prepositional phrases.[6]

2.2 Movement of Noun Phrases to Subject Position

We have encoded unaccusativity in the schema. If there is no subject of the verb phrase, then the object moves to the subject position in the reading of the sentence. We also have encoded the other thematic roles using prepositions, both overt and covert. If there is a prepositional phrase and the preposition is overt, then the same rule of the direct object moving to the subject position applies here also. The movement of noun phrases to the subject position when there exist in the sentence prepositional phrases with hidden prepositions will be seen later. We restate this procedurally in (1):

(1) a. If there is a subject of the verb phrase, it becomes the subject of the sentence.
b. If there is no subject of the verb phrase and there is no hidden preposition (except around), then the object of the verb phrase becomes the subject of the sentence.
c. Otherwise, the object of a hidden preposition becomes the subject of the sentence.

The below example (2), example (12) of the unaccusativity hypothesis in section 1.1.2.1 of EVCA, shows a sentence with no hidden prepositions with this sort of argument structure.[7]

[6] See Doug Jones and Anand Radhakrishnan's "Transitive Prepositions as Obligatory Adjuncts" in this collection for a discussion.

[7] In this representation of argument structure, vp means verb phrase, and an e means that this element is empty. Our current representation of (2.a.) is [np,v,np], which is parsed

(2)

SENTENCE	ARGUMENT STRUCTUREMatrix
a. Janet broke the cup.	vp(v, np, np)
b. The cup broke.	vp(v, e, np)

2.3 Prepositions Yield Thematic Roles

Most of the alternations in EVCA hinge on the placement of prepositions and their objects. Prepositions in the majority of the alternations imply binary relationships, such as figure and ground or material and artifact, between noun phrases. These relationships can be counted and numbered, so a preposition is given a reading number telling what kind of noun phrase its object should be. Since there is a limited number of prepositions and a small number of readings for each preposition, the preposition is a good location for information to be stored about noun phrases and relationships. We have encoded nine readings of one preposition and only one of another, but these readings are now easily understandable. This encoding of information into the prepositions also helped to reduce the number of schema elements. In the example (3) below, the preposition to of sentence (3.a.) indicates that its object will be the indirect object of that sentence. In sentence (3.b.), the indirect object has moved to the position before the direct object and has lost the preposition *to*.

(3) a. Bill sold a car to Tom.
    b. Bill sold Tom a car.

It seems in (3.b) that the *to* is still understood. We could say that this *to* is a hidden preposition. An alternation with noun phrases other than agent,

---

as vp(v,np,np), from which we get its thematic assignment. The vp subject becomes then sentence agent, and the vp object becomes the sentence theme. The verb phrase subject in vp(v,np,np) is the first np, and the verb phrase object is the second np.

object, bodyparts, animals, and cognates depended only on the prepositions and their objects. Using this idea of hidden prepositions, all noun phrases, other than those just mentioned, not in prepositional phrases could now be placed in prepositional phrases with hidden prepositions. Each hidden preposition would be denoted by _0 appended to the end of it in Prolog notation. These prepositions, like the overt prepositions, would have reading numbers indicating the kinds of noun phrases to be their objects.

(4) a. Jack sprayed paint on the wall.
[np,v,np(figure),[p(on),np(ground)]]
[np,v,[p(with_0,4),np],[p(on,5),np]]
b. Jack sprayed the wall with paint.
[np,v,np(ground),[p(with),np(figure)]]
[np,v,[p(on_0,1),np],[p(with,7),np]]

We are considering using instead of two prepositional phrases, one of which has a hidden preposition, a representation that allows a preposition to have a subject as well as an object. In changing this representation, we avoid can avoid the question of obligatory adjunction of prepositional phrases or adverbs.[8]

When there is a hidden preposition in a sentence with no subject, the rule for movement to the subject position usually means the object of the hidden preposition, not the direct object, moving. The one instance found in this study in which the direct object, rather than the object of the hidden preposition, moves is the case of a hidden around. For the other prepositions studied, *with*, *to*, *in*, *for*, *from*, *into*, and *of*, it does not matter whether there exists in the sentence a direct object, since the object of the hidden preposition always moves. The thematic roles in example (5) are derived as described in (1).

(5) 1. a. Mary rotates the top.

[8] See Doug Jones and Anand Radhakrishnan's "Transitive Prepositions as Obligatory Adjuncts" in this collection for a discussion.

b. The top rotates.
c. The top rotates around its axis.
d. *The top rotates its axis.

2. a. I cut the bread with this knife.
   b. This knife cut the bread.
   c. This knife doesn't cut.

Another possible modification that we have considered is to disambiguate the different meanings of a single preposition. In the below example, the around in (6.a.) means "in a path avoiding (an obstacle)" while the around in (6.b.) could be replaced with the phrase "around in". Then the length of the list of prepositions would lengthen, and the reading numbers would separate among the different meanings of the prepositions.

(6) a. The boat floats around the buoy.
    b. The boat floats around the harbor.

The parses have been presented in the form of lists as can be seen in the above example (4). Recently we have been considering changing this list form to the form following xbar theory. X-bar theory parses sentences or sentence parts into a tree structure as shown below. The XP can be any kind of phrase: inflection phrase, verb phrase, prepositional phrase, or other. The XP is made of three parts: the head X^0, a specifier, and a complement. The head is the verb or preposition. The specifier is the subject of the phrase, and the complement is the object of the phrase. The specifier or the complement can themselves be phrases, other XPs. Sometimes the specifier or the complement may not exist. We have noted these with an *e*.

```
      XP
      /\
     /  \
    /    \
   /     /\
  /     /  \
Spec  Head  Compl
```

Doug Jones wrote a routine that will change many of these to the X-bar form. Only those parses that passed this test were kept to be added to EVCA WordNet. Those that passed were 91this routine adjunction of prepositional phrases and adverbs to the main clause is allowed. This allows a structure with more than the usual three parts. Pure X-bar theory without adjunction to XP does not allow this. (As a first approximation, we did not consider adjundtion to XP). If we had not allowed adjunction to XP, then the coverage of the sentences would be 60this does not include most two-argument prepositional phrases. These two- argument prepositional phrases would be two prepositional phrases, the first of which had a hidden preposition. If two-place prepositional phrases were allowed, then the coverage would be 83

2.4 Toyworld

In several cases, a single specific noun phrase could be used throughout all the alternations to replace all or most instances of a certain kind of noun phrase in the schema. For example, almost all verbs could take a human subject, *Mary*. The subjects for the others that could not take a human were given separately. Some verbs can take a kind of subject other than a human subject. WordNet gives two separate sample sentences for these verbs, using somebody and something. In adding real sentences to WordNet, only one sentence per schema per sense was created, so only the human subject was chosen if it made sense. See Figure **??** in Appendix C for an example.

In order to come up with the rest of the good noun phrases for the sentences, it was necessary to find which properties a noun phrase needed to make sense in a sentence with a particular verb. So a verb possesses a list of properties for each noun phrase in its alternation. These property lists would include such general descriptions as thing, animal, person, solid, liquid, and abstract, as well as some more specific qualities like *texture:springy*, *shape:axial*, *feathered*, and *physical_property:flammable*. Sometimes a specific noun would be included if the verb restricted nouns for one noun phrase to one of a very few choices. In the cases of noun phrases other than subjects also, many verbs can take either a human or an inanimate noun. WordNet gives these, too, as *somebody* and *something*. Since only one sentence was given from EVCA, the more appropriate noun was chosen, and it would be given that property list. See Figure **??** in Appendix C for an example.

In most cases, verbs grouped themselves together by requiring, at least most of the time, the same kind of noun for a particular noun phrase. For example, very many verbs needed solid direct objects in most ways that they were used. For example, sense 3 of *move* takes a solid object, as do *drop*, *hit*, *put*, and *shellac*. *Move* is a fairly general verb that can be done to a solid object. So we say that *drop*, *hit*, *put*, *shellac*, and many others inherit the property list of the verb *move*. We use this idea of linking through inheritances to propagate the property solid through all the verbs that need a direct object that is solid. These verbs that inherit the property lists of another verb are obviously not all synonyms. We simply used this technique of inheritances to show that they share the relevant properties of the noun phrase. In cases where there are noun phrases that pair themselves together almost all of the time, such as figure and ground or material and artifact, we represented the inheritances slightly differently. Sometimes two verbs could take the same sort of ground, but not the same type of figure. We decided that a verb inherits from another verb if and only if it inherits the properties of both noun phrases.

What is the shape of our lexical inheritance system? We call the general verb a parent. A verb that inherits its properties from another verb we call a child. There were always many more children than parents. The object of the verb phrase was the most common of all the noun phrases that were given property lists. In this case, there were 127 parents and 2063 children, giving a ratio of more than 16 children to one parent; the distribution is relatively flat. For an instrument noun phrase, there were twelve parents and 112 children. In the case of figure and ground, where the two noun phrases usually pair themselves together, the number of parents of figure, though their property lists were not all unique, equaled eighty and the number of parents of ground, though their property lists also were not all unique, equaled eighty-one. This difference is due to the fact that in a few cases, figure and ground noun phrases did not appear together in the sentence. Their children, inheriting the pair of property lists from figure and from ground, totalled 647. Other noun phrases did not appear in so many of the sentence alternations.

The lists of properties ranged from very general, e.g. thing, to very specific, e.g. egg. The property lists of direct objects included 127 properties together. For figure, there were 54 properties, and 28 for ground. Other noun phrases had few total properties among all the property lists. Some of these surely overlap among different noun phrases. The instances of real nouns with these properties are also divided by the types of noun phrases. One improvement would be to keep the division by noun phrase type in the assignment of property lists to noun phrases of verbs but to remove the division in the real instances of noun phrases, since there is plenty of overlapping in the property lists among these types of noun phrases. Some property lists include properties such as shape:round, texture:flexible, or texture:springy. These give a feature that the noun phrase depends on and then what kind of quality that feature must have. In the direct objects, there are five different features and 37 qualities among these five features that the feature may have.

These property lists were built solely by intuition. In many cases, we may be able to derive the properties from the verb. As an example, the property list of the direct object of *bounce* is *[thing,solid,texture:springy]*. But *springy* means simply that this object can bounce,or is bounceable. We are considering creating an operator that would derive this property from the verb. More specifically, *able(X)*, where *X* is the verb would create the property *X+able*. So *able(bounce)* would produce *bounce+able*, which could replace *texture:springy*.

Sometimes a verb can take only a very limited kind of direct object or other noun phrase. For example, *hatch* usually refers to a bird coming out of its egg. "To hatch a plan" would be a figurative use. Here we are considering defining a meta-operator, *refer(X)*, that would give this specific noun, egg, as the usual direct object of *hatch*. So *refer(hatch)* would give *egg* as the direct object.

We have encoded all of the lexical features into the noun phrases in their property lists so far. There is some other information that cannot be encoded in the noun phrases. We are considering adding this information into other parts of the sentence. For example, *drop* means "to go straight down" in simple terms. This means that the path of an object dropping has to be straight. Sentence (7.a.) would be a grammatical sentence, but sentence

(7.b.) would be ungrammatical. The preposition around means that the path cannot be straight. The two paths do not agree, so the sentence is a bad one.

(7) a. The pine cone dropped past the tree branch.
    b. *The pine cone dropped around the tree.

2.5 The Sentence Generator

I wrote the program to generate these sentences in Prolog. Using this program, it is possible to create 4 levels of description of a sentence for WordNet to read. The simplest give solely the generated sentence. Another produces, in addition to the sentence, the thematic roles of many noun phrases. The third gives the sentence, the thematic roles, and the property lists of many noun phrases. The fourth gives the sentence, the thematic roles, and the property lists. The fourth also gives a few of the spray/load paraphrases discussed below. On a Sparc 10, it takes about 8 minutes to create a Prolog file of these sentences with the thematic roles and property lists and then about 15 minutes to read this file into Prolog and create three other files for EVCA WordNet to read. See examples of EVCA WordNet sessions in Appendix C.

2.6 Spray/Load paraphrases

In an earlier study of the Spray/Load alternation in verbs, we noticed that many of the alternations could be paraphrased as described in (8) using light verbs which do not contain much meaning. Any that pass both tests are in the Spray/Load Alternating class. Any that pass just the first test belong to the Non-Alternating Locative Preposition Only class. Any that pass the second are in the Non-Alternating With Only class. These templates would provide aid to someone learning the language. They could also helpful in studying the semantics of the verbs in these classes. See Figure **??** in Appendix C for a session of WordNet with these paraphrases.

(8) 1. [np(subject),verb,np(figure),prep,np(ground)] (prep is not with)
becomes
np(subject) causes np(figure) to go prep np(ground) by verb+ing np(figure).
2. [np(subject),verb,np(ground),prep(with),np(figure)]
becomes
np(subject) causes np(ground) to have np(figure) by verb+ing np(ground).

The below example (9) shows several Alternating verbs that do follow these templates. In more than a few cases, Anand Radhakrishnan noticed that these templates do not always hold true within each class of alternating/non-alternating. The verbs have divided themselves classes into subclasses which behave similarly in these templates and in another suggested alternation. While these are all Spray/Load verbs, example (9) shows that spray is correct in a both paraphrases, dust in only one paraphrase, and pile in the other. See "Transitive Prepositions as Obligatory Adjuncts" in this collection for a more complete discussion.

(9) 1. a. Mary sprayed the paint on the wall.
Mary caused the paint to go on the wall by spraying the paint.
b. Mary sprayed the wall with paint.
Mary caused the wall to have paint by spraying the wall.
2. a. I dust the cloth over the shelf.
* I cause the cloth to go over the shelf by dusting the cloth.
b. I dust the shelf with the cloth.
I cause the shelf to have the cloth by dusting the shelf.
3. a. I pile the books onto the shelf.
I pile the shelf with the books.
b. I cause the books to go onto the shelf by piling the books.
*I cause the shelf to have the books by piling the shelf.

See examples of EVCA WordNet sessions in Appendix C. Figures (1) (a-d) shows the four-level contrast in WordNet sessions. Figure **??** shows the hand-inserted paraphrases of the verb spray in addition to the generated sample sentences.

3 Enhancing EVCA with WordNet Wordsenses

3.1 A Survey of the Word Senses in WordNet and EVCA

An important addition to EVCA WordNet was the differentiation of the word senses of each verb. WordNet gives one or more word sense numbers per verb. The number of word senses is the number of different meanings of the verb. This was an important enhancement since quite a few EVCA verbs appeared in more than one class, or dataset. Often the appearance of a verb in more than one class meant that there were different senses. Now there should be no confusion as to the meaning or the alternation.

The determining of the word sense of a verb was not performed mechanically. It required just by looking at the alternation, the WordNet synonyms and sample sentences, and the other verbs in that verb class, or dataset. A verb with a WordNet word sense became noted as the verb with a hyphen and the sense number after it. Sense 2 of bake would become bake-2.

WordNet contains 11340 verbs in version 1.3. These verbs correspond to 18749 word senses. So the ratio of senses to verbs is 1.65. In EVCA, there are 2600 verbs with 3017 senses, therefore a ratio of 1.16. Four hundred forty-seven of these verbs are not in WordNet. Three hundred seventy-three of these 446 verbs do not appear in WordNet at all. The other 73 verbs are in WordNet but do not have the correct sense. For example, *bat* appears in WordNet with two senses:

2 senses of bat

Sense 1
bat, strike with the bat – (in baseball)
  *¿ Somebody —-s something
Sense 2
bat, clobber, drub, thrash, lick, beat –(in sports)

*¿ Somebody —-s something
*¿ Somebody —-s somebody

Neither of these is the correct sense when used in the phrase "batting one's eyes." The 74 verbs with missing senses correspond to 76 missing senses, since two of these 74 verbs are missing two senses each. Without these 446 verbs, EVCA and WordNet overlap with 2571 word senses, which is 14WordNet's word senses. See Figure **??** in Appendix D for a distribution of these WordNet senses in EVCA.

WordNet has many verbs that constitute several words but are still considered a single verb. For example, abstain appears in WordNet with 2 senses. But also included in WordNet are *abstain_from*, *abstain_from_food*, *abstain_from_certain_foods*, *abstain_from_voting*, and *promise_to_abstain_from*, all with very similar or specific meanings. WordNet makes very fine-grained distinctions such as those among *write_with_a_pencil*, *write_with_chalk*, and *write_with_a_crayon*. WordNet has 1920 multi-word entries and EVCA has 8 multi-word entries. Without these multi-word verbs, WordNet has only 9420 single-word verb entries, and EVCA has 2592 single-word verb entries, 2206 of which are in WordNet. These 2206 make 23single-word entries. These figures do not count the word senses of the single-word entries. Since most of these multi-word entries have just one word sense, the ratio of WordNet's senses to verbs would increase slightly if these multi-word entries were not counted.

In Levin's study of verb class alternations, she concludes that a sense of a verb is derived from another sense because of these alternations. In example (10) below, Levin's example (12) of section 1.1.2.1, the two usages of break are considered the same word sense. See also Figure **??** in Appendix C. The second is derived from the first by the unaccusative hypothesis.

(10) a. Janet broke the cup. [no,v,np]
b. The cup broke. [v,np]

In WordNet, however, the verbs in (10.a.) and (10.b.) are considered to have different word senses. As a rule in annotating the verbs with sense numbers, the causative, if it existed in WordNet, was chosen to be the correct sense. If the EVCA example sentences of the alternations had been split between the two senses in WordNet, then the alternation would not have been clear. It was important to keep these alternations together. In this project we are trying to understand the relationships among the senses of verbs. If all these word senses had also been counted, then the percent coverage of WordNet would have increased noticeably. Also, the gap between EVCA ratio of 1.16 of senses to verbs and the WordNet ratio of 1.65 senses to verbs would have narrowed.

Sometimes more than one word sense of a verb fits the alternation. Instead of adding all the possible senses that could fit into the alternation, only the most familiar or most general of the senses was chosen for the correct number for now. These other possible senses have similar or figurative meanings. See Figure **??** in Appendix C for an example. In this example of the verb *burn*, it is obvious that more senses than sense 2 could fit the alternation of example (11). Of course sense 3 is derived from sense 2, as is sense 6 from sense 5 and sense 9 from sense 8. Example sentences created for EVCA WordNet in sense 2 are below. Example sentences of this alternation could also be made from sense numbers 4, 5, 7, 8, and 10, as in the below example (11).

(11) [np,v,np]
[v,np]

2. a. Mary burns the leaves.
   b. The leaves burn.

4. a. Mary is burning the witch.
   b. The witch is burning.
5. a. The pepper is burning my eye.
   b. My eye is burning.
7. a. Mary is burning the building.
   b. The building is burning.
8. a. Mary is burning the log.
   b. The log is burning.

10. a. The acid is burning my skin.
    b. My skin is burning.

3.2 Semantic Classes Missing from WordNet

Four hundred forty-six verb senses of EVCA verbs are not in WordNet. This is 15entries are not limited to single root forms. The 11,340 entries reduce to 9420 if the multi-word verbs, such as *abstain_from_eating_certain_foods*, *abstain_from_voting, write_with_a_crayon*, and *write_with_a_pencil* are removed. Since these verbs have more than the single root form, we would expect just as many other verbs with more than the root to appear in WordNet. We notice, however, that many verbs produced from derivational morphology such as that involving *de-* and *un-* are missing. For example, the EVCA verb class "Debone Verbs" from section 10.8 would virtually disappear if we remove the EVCA verbs that have no sense in WordNet–down to 16in WordNet are *declaw*, *dehair*, and *deink*; other verbs with the prefix *un-* that are not included in WordNet are *unclamp*, *unlace*, and *unzip*. There are very few verbs with these prefixes in WordNet. Several other classes would completely disappear, but these had very few members (1-3) originally. See Figure **??** in Appendix D for the percentage distribution of WordNet verbs through EVCA verb classes.

4 Other Suggestions for EVCA WordNet

If more verbs were to be added to EVCA WordNet in an existing verb class, it would be necessary to determine what properties were necessary for any of the noun phrases. If another verb with the same properties for the same noun phrase already existed, then the new verb would have to be added as a child of the other verb. If no verb with the same property list for the same noun phrase existed, then it would be necessary to find an good noun phrase having this property list.

In generating these sentences, I added only a *+s*, *+ed*, *+en*, or *+ing* to show the tenses. Our group is considering writing a morphological analyzer to change these endings to the correct verb form. This analyzer would be added on after these sentences were generated. For example, *give+ed* would change to *gave*x.

A WordNet Verb Frames

1 Something ----s
2 Somebody ----s
3 It is ----ing
4 Something is ----ing PP
5 Something ----s something Adjective/Noun
6 Something ----s Adjective/Noun
7 Somebody ----s Adjective
8 Somebody ----s something
9 Somebody ----s somebody
10 Something ----s somebody
11 Something ----s something
12 Something ----s to somebody
13 Somebody ----s on something
14 Somebody ----s somebody something
15 Somebody ----s something to somebody
16 Somebody ----s something from somebody
17 Somebody ----s somebody with something
18 Somebody ----s somebody of something
19 Somebody ----s something on somebody
20 Somebody ----s somebody PP
21 Somebody ----s something PP
22 Somebody ----s PP
23 Somebody's (body part) ----s
24 Somebody ----s somebody to INFINITIVE
25 Somebody ----s somebody INFINITIVE
26 Somebody ----s that CLAUSE
27 Somebody ----s to somebody
28 Somebody ----s to INFINITIVE
29 Somebody ----s whether INFINITIVE
30 Somebody ----s somebody into V-ing something
31 Somebody ----s something with something
32 Somebody ----s INFINITIVE
33 Somebody ----s VERB-ing
34 It ----s that CLAUSE

B EVCA Verb Classes, or Datasets, in Prolog

In the Prolog representation below, 12:a and 12:b mean examples (12.a.) and (12.b.) in EVCA. The letter *s* means sentence, and the ones or zeroes mean that this pattern is grammatical or ungrammatical for this particular verb class. The example sentences are given with their parses just below.

```
evca_dataset(2,
     [coil-3, revolve-2, rotate-1, spin-1, turn-2, twirl-1, twist-5,
     whirl-1, wind-3],
     [pattern(7:ii,'Motion Around an Axis',
          [
           eg(12:a,s,1,
              'Janet broke the cup.',
              [np,v,np]),
           eg(12:b,s,1,
              'The cup broke.',
              [v,np])]),
      pattern(105:ii,'Verbs of Motion Around an Axis',
          [
           eg(106:a,s,1,
              'The spaceship revolves around the earth.',
              [v,np,[p(around,1),np]]),
           eg(106:b,s,0,
              'The spaceship revolves the earth.',
              [v,np,[p(around_0,1),np]])])]).


evca_dataset(101,
     [brush-3, cram-1, crowd-1, cultivate-2, dab-1, daub-1, drape-2,
     drizzle-2, dust-1, hang-4, heap-1, inject-6, jam-3, load-2, mound,
     pack-4, pile-2, plant-3, plaster-3, pump-2, rub-3, scatter-3, seed,
     settle-3, sew-2, shower-2, slather, smear-3, smudge-1, sow,
     spatter-3, splash-1, splatter-1, spray-1, spread-3, sprinkle-2,
     spritz, squirt-1, stack-2, stick-1, stock-1, strew-1, string-7,
     stuff-1, swab-2, wrap-1],
     [pattern(124,'Spray/Load Verbs',
          [
           eg(125:a,s,1,
```

```
                'Jack sprayed paint on the wall.',
                [np,v,[p(with_0,4),np],[p(on,5),np]]),
             eg(125:b,s,1,
                'Jack sprayed the wall with paint.',
                [np,v,[p(on_0,1),np],[p(with,7),np]])])]).


evca_dataset(126,
     [alter-4, change-4, convert-3, metamorphose-1, transform-1,
     transmute-1, turn-10],
     [pattern(156,'Turn Verbs',
          [
           eg(157:a,s,0,
                'He turned from a prince.',
                [v,np,[p(from,6),np]]),
           eg(157:b,s,1,
                'He turned into a frog.',
                [v,np,[p(into,2),np]])]),
      pattern(158,'Turn Verbs',
          [
           eg(159:a,s,1,
                'The witch turned him into a frog.',
                [np,v,np,[p(into,2),np]]),
           eg(159:b,s,1,
                'The witch turned him from a prince into a frog.',
                [np,v,np,[p(from,6),np],[p(into,2),np]])]),
      pattern(150:b,'Turn Verbs',
          [
           eg(151:a,s,1,
                'I kneaded the dough into a loaf.',
                [np,v,np,[p(into,2),np]]),
           eg(151:b,s,0,
                'I kneaded a loaf from the dough.',
                [np,v,[p(into_0,2),np],[p(from,1),np]]),
           eg(152:a,s,1,
                'The witch turned him into a frog.',
                [np,v,np,[p(into,2),np]]),
           eg(152:b,s,0,
                'The witch turned him from a prince.',
                [np,v,np,[p(from,6),np]])])]).
```