

# Extended Lexical-Semantic Classification of English Verbs

**Anna Korhonen and Ted Briscoe**

University of Cambridge, Computer Laboratory  
15 JJ Thomson Avenue, Cambridge CB3 0FD, UK  
alk23@cl.cam.ac.uk, ejb@cl.cam.ac.uk

## Abstract

Lexical-semantic verb classifications have proved useful in supporting various natural language processing (NLP) tasks. The largest and the most widely deployed classification in English is Levin's (1993) taxonomy of verbs and their classes. While this resource is attractive in being extensive enough for some NLP use, it is not comprehensive. In this paper, we present a substantial extension to Levin's taxonomy which incorporates 57 novel classes for verbs not covered (comprehensively) by Levin. We also introduce 106 novel diathesis alternations, created as a side product of constructing the new classes. We demonstrate the utility of our novel classes by using them to support automatic subcategorization acquisition and show that the resulting extended classification has extensive coverage over the English verb lexicon.

## 1 Introduction

Lexical-semantic classes which aim to capture the close relationship between the syntax and semantics of verbs have attracted considerable interest in both linguistics and computational linguistics (e.g. (Pinker, 1989; Jackendoff, 1990; Levin, 1993; Dorr, 1997; Dang et al., 1998; Merlo and Stevenson, 2001)). Such classes can capture generalizations over a range of (cross-)linguistic properties, and can therefore be used as a valuable means of reducing redundancy in the lexicon and for filling gaps in lexical knowledge.

Verb classes have proved useful in various (multilingual) natural language processing (NLP) tasks and applications, such as computational lexicography (Kipper et al., 2000), language generation (Stede, 1998), machine translation (Dorr, 1997), word sense disambigua-

tion (Prescher et al., 2000), document classification (Klavans and Kan, 1998), and subcategorization acquisition (Korhonen, 2002). Fundamentally, such classes define the mapping from surface realization of arguments to predicate-argument structure and are therefore a critical component of any NLP system which needs to recover predicate-argument structure. In many operational contexts, lexical information must be acquired from small application- and/or domain-specific corpora. The predictive power of classes can help compensate for lack of sufficient data fully exemplifying the behaviour of relevant words, through use of back-off smoothing or similar techniques.

Although several classifications are now available for English verbs (e.g. (Pinker, 1989; Jackendoff, 1990; Levin, 1993)), they are all restricted to certain class types and many of them have few exemplars with each class. For example, the largest and the most widely deployed classification in English, Levin's (1993) taxonomy, mainly deals with verbs taking noun and prepositional phrase complements, and does not provide large numbers of exemplars of the classes. The fact that no comprehensive classification is available limits the usefulness of the classes for practical NLP.

Some experiments have been reported recently which indicate that it should be possible, in the future, to automatically supplement extant classifications with novel verb classes and member verbs from corpus data (Brew and Schulte im Walde, 2002; Merlo and Stevenson, 2001; Korhonen et al., 2003). While the automatic approach will avoid the expensive overhead of manual classification, the very development of the technology capable of large-scale automatic classification will require access to a target classification and gold standard exemplification of it more extensive than that available currently.

In this paper, we address these problems by introducing a substantial extension to Levin's classification which incorporates 57 novel classes for verbs not covered (com-

prehensively) by Levin. These classes, many of them drawn initially from linguistic resources, were created semi-automatically by looking for diathesis alternations shared by candidate verbs. 106 new alternations not covered by Levin were identified for this work. We demonstrate the usefulness of our novel classes by using them to improve the performance of our extant subcategorization acquisition system. We show that the resulting extended classification has good coverage over the English verb lexicon. Discussion is provided on how the classification could be further refined and extended in the future, and integrated as part of Levin's extant taxonomy.

We discuss Levin's classification and its extensions in section 2. Section 3 describes the process of creating the new verb classes. Section 4 reports the experimental evaluation and section 5 discusses further work. Conclusions are drawn in section 6.

## 2 Levin's Classification

Levin's classification (Levin, 1993) provides a summary of the variety of theoretical research done on lexical-semantic verb classification over the past decades. In this classification, verbs which display the same or similar set of diathesis alternations in the realization of their argument structure are assumed to share certain meaning components and are organized into a semantically coherent class. Although alternations are chosen as the primary means for identifying verb classes, additional properties related to subcategorization, morphology and extended meanings of verbs are taken into account as well.

For instance, the Levin class of "Break Verbs" (class 45.1), which refers to actions that bring about a change in the material integrity of some entity, is characterized by its participation (1-3) or non-participation (4-6) in the following alternations and other constructions (7-8):

1. **Causative/inchoative alternation:**  
*Tony broke the window ↔ The window broke*
2. **Middle alternation:**  
*Tony broke the window ↔ The window broke easily*
3. **Instrument subject alternation:**  
*Tony broke the window with the hammer ↔ The hammer broke the window*
4. **\*With/against alternation:**  
*Tony broke the cup against the wall ↔ \*Tony broke the wall with the cup*
5. **\*Conative alternation:**  
*Tony broke the window ↔ \*Tony broke at the window*
6. **\*Body-Part possessor ascension alternation:**  
*\*Tony broke herself on the arm ↔ Tony broke her arm*
7. **Unintentional interpretation available (some verbs):**  
Reflexive object: *\*Tony broke himself*  
Body-part object: *Tony broke his finger*

## 8. Resultative phrase:

*Tony broke the piggy bank open, Tony broke the glass to pieces*

Levin's taxonomy provides a classification of 3,024 verbs (4,186 senses) into 48 broad and 192 fine-grained classes according to their participation in 79 alternations involving NP and PP complements.

Some extensions have recently been proposed to this resource. Dang et al. (1998) have supplemented the taxonomy with intersective classes: special classes for verbs which share membership of more than one Levin class because of regular polysemy. Bonnie Dorr (University of Maryland) has provided a reformulated and extended version of Levin's classification in her LCS database (<http://www.umiacs.umd.edu/~bonnie/verbs-English.lcs>). This resource groups 4,432 verbs (11,000 senses) into 466 Levin-based and 26 novel classes. The latter are Levin classes refined according to verbal telicity patterns (Olsen et al., 1997), while the former are additional classes for non-Levin verbs which do not fall into any of the Levin classes due to their distinctive syntactic behaviour (Dorr, 1997).

As a result of this work, the taxonomy has gained considerably in depth, but not to the same extent in breadth. Verbs taking ADJP, ADVP, ADL, particle, predicative, control and sentential complements are still largely excluded, except where they show interesting behaviour with respect to NP and PP complementation. As many of these verbs are highly frequent in language, NLP applications utilizing lexical-semantic classes would benefit greatly from a linguistic resource which provides adequate classification of their senses. When extending Levin's classification with new classes, we particularly focussed on these verbs.

## 3 Creating Novel Classes

Levin's original taxonomy was created by

1. selecting a set of diathesis alternations from linguistic resources,
2. classifying a large number of verbs according to their participation in these alternations,
3. grouping the verbs into semantic classes based on their participation in sets of alternations.

We adopted a different, faster approach. This involved

1. composing a set of diathesis alternations for verbs not covered comprehensively by Levin,
2. selecting a set of candidate lexical-semantic classes for these verbs from linguistic resources,

3. examining whether (sub)sets of verbs in each candidate class could be related to each other via alternations and thus warrant creation of a new class.

In what follows, we will describe these steps in detail.

### 3.1 Novel Diathesis Alternations

When constructing novel diathesis alternations, we took as a starting point the subcategorization classification of Briscoe (2000). This fairly comprehensive classification incorporates 163 different subcategorization frames (SCFs), a superset of those listed in the ANLT (Boguraev et al., 1987) and COMLEX Syntax dictionaries (Grishman et al., 1994). The SCFs define mappings from surface arguments to predicate-argument structure for bounded dependency constructions, but abstract over specific particles and prepositions, as these can be trivially instantiated when the a frame is associated with a specific verb. As most diathesis alternations are only semi-predictable on a verb-by-verb basis, a distinct SCF is defined for every such construction, and thus all alternations can be represented as mappings between such SCFs.

We considered possible alternations between pairs of SCFs in this classification, focusing in particular on those SCFs not covered by Levin. The identification of alternations was done manually, using criteria similar to Levin's: the SCFs alternating should preserve the sense in question, or modify it systematically.

106 new alternations were discovered using this method and grouped into different, partly overlapping categories. Table 1 shows some example alternations and their corresponding categories. The alternating patterns are indicated using an arrow ( $\leftrightarrow$ ). The SCFs are marked using number codes whose detailed description can be found in (Briscoe, 2000) (e.g. SCF 53. refers to the COMLEX subcategorization class NP-TO-INF-OC).

### 3.2 Candidate Lexical-Semantic Classes

Starting off from set of candidate classes accelerated the work considerably as it enabled building on extant linguistic research. Although a number of studies are available on verb classes not covered by Levin, many of these assume a classification system completely different to that of Levin's, and/or incorporate sense distinctions too fine-grained for easy integrations with Levin's classification. We therefore restricted our scope to a few classifications of a suitable style and granularity:

#### 3.2.1 The LCS Database

The LCS database includes 26 classes for verbs which could not be mapped into any of the Levin classes due to their distinctive syntactic behaviour. These classes were originally created by an automatic verb classification algorithm described in (Dorr, 1997). Although they appear semantically meaningful, their syntactic-semantic

properties have not been systematically studied in terms of diathesis alternations, and therefore re-examination is warranted.

#### 3.2.2 Rudanko's Classification

Rudanko (1996, 2000) provides a semantically motivated classification for verbs taking various types of sentential complements (including predicative and control constructions). His relatively fine-grained classes, organized into sets of independent taxonomies, have been created in a manner similar to Levin's. We took 43 of Rudanko's verb classes for consideration.

#### 3.2.3 Sager's Classification

Sager (1981) presents a small classification consisting of 13 classes, which groups verbs (mostly) on the basis of their syntactic alternations. While semantic properties are largely ignored, many of the classes appear distinctive also in terms of semantics.

#### 3.2.4 Levin's Classification

At least 20 (broad) Levin classes involve verb senses which take sentential complements. Because full treatment of these senses requires considering sentential complementation, we re-evaluated these classes using our method.

### 3.3 Method for Creating Classes

Each candidate class was evaluated as follows:

1. We extracted from its class description (where one was available) and/or from the COMLEX Syntax dictionary (Grishman et al., 1994) all the SCFs taken by its member verbs.
2. We extracted from Levin's taxonomy and from our novel list of 106 alternations all the alternations where these SCFs were involved.
3. Where one or several alternations were found which captured the sense in question, and where the minimum of two member verbs were identified, a new verb class was created.

Steps 1-2 were done automatically and step 3 manually. Identifying relevant alternations helped to identify additional SCFs, which in turn often led to the discovery of additional alternations. The SCFs and alternations discovered in this way were used to create the syntactic-semantic description of each novel class.

For those candidate classes which had an insufficient number of member verbs, new members were searched for in WordNet (Miller, 1990). Although WordNet classifies verbs on a purely semantic basis, the syntactic regularities studied by Levin are to some extent reflected

Category	Example Alternations	Alternating SCFs
Equi	<i>I advised Mary to go</i> ↔ <i>I advised Mary</i> <i>He helped her bake the cake</i> ↔ <i>He helped bake the cake</i>	53 ↔ 24 33 ↔ 142
Raising	<i>Julie strikes me as foolish</i> ↔ <i>Julie strikes me as a fool</i> <i>He appeared to her to be ill</i> ↔ <i>It appeared to her that he was ill</i>	143 ↔ 29 99 ↔ 12
Category switches	<i>He failed in attempting to climb</i> ↔ <i>He failed in the climb</i> <i>I promised Mary to go</i> ↔ <i>I promised Mary that I will go</i>	63 ↔ 87 54 ↔ 52
PP deletion	<i>Phil explained to him how to do it</i> ↔ <i>Phil explained how to do it</i> <i>He contracted with him for the man to go</i> ↔ <i>He contracted for the man to go</i>	90 ↔ 17 88 ↔ 15
P/C deletion	<i>I prefer for her to do it</i> ↔ <i>I prefer her to do it</i> <i>They asked about what to do</i> ↔ <i>They asked what to do</i>	15 ↔ 53 73 ↔ 116

Table 1: Examples of new alternations

by semantic relatedness as it is represented by WordNet’s particular structure (e.g. (Fellbaum, 1999)). New member verbs were frequently found among the synonyms, troponyms, hypernyms, coordinate terms and/or antonyms of the extant member verbs.

For example, using this method, we gave the following description to one of the candidate classes of Rudanko (1996), which he describes syntactically with the single SCF 63 (see the below list) and semantically by stating that verbs in this class (e.g. *succeed, manage, fail*) have approximate meaning<sup>1</sup> “perform the act of” or “carry out the activity of”:

#### 20. SUCCEED VERBS

- SCF 22: *John succeeded*  
 SCF 87: *John succeeded in the climb*  
 SCF 63: *John succeeded in attempting the climb*  
 SCF 112: *John succeeded to climb*

**Alternating SCFs:** 22 ↔ 87, 87 ↔ 63, 22 ↔ 112

Some of the candidate classes, particularly those of Rudanko, proved too fine-grained to be helpful for a Levin type of classification, and were either combined with other classes or excluded from consideration. Some other classes, particularly the large ones in the LCS database, proved too coarse-grained after our method was applied, and were split down to subclasses.

For example, the LCS class of *Coerce Verbs* (002) was divided into four subclasses according to the particular syntactic-semantic properties of the subsets of its member verbs. One of these subclasses was created for verbs such as *force, induce, and seduce*, which share the ap-

<sup>1</sup>Rudanko does not assign unique labels to his classes, and the descriptions he gives - when taken out of the context - cannot be used to uniquely identify the meaning involved in a specific class. For details of this class, see his description in (Rudanko, 1996) page 28.

proximate meaning of “urge or force (a person) to an action”. The sense gives rise to object equi SCFs and alternations:

#### 2. FORCE VERBS

- SCF 24: *John forced him*  
 SCF 40: *John forced him into coming*  
 SCF 49: *John forced him into it*  
 SCF 53: *John forced him to come*

**Alternating SCFs:** 24 ↔ 53, 40 ↔ 49, 49 ↔ 24

Another subclass was created for verbs such as *order* and *require*, which share the approximate meaning of “direct somebody to do something”. These verbs take object raising SCFs and alternations:

#### 3. ORDER VERBS

- SCF 57: *John ordered him to be nice*  
 SCF 104: *John ordered that he should be nice*  
 SCF 106: *John ordered that he be nice*

**Alternating SCFs:** 57 ↔ 104, 104 ↔ 106

New subclasses were also created for those Levin classes which did not adequately account for the variation among their member verbs. For example, a new class was created for those 37. Verbs of Communication which have an approximate meaning of “make a proposal” (e.g. *suggest, recommend, propose*). These verbs take a rather distinct set of SCFs and alternations, which differ from those taken by other communication verbs. This class is somewhat similar in meaning to Levin’s 37.9 *Advise Verbs*. In fact, a subset of the verbs in 37.9 (e.g. *advise, instruct*) participate in alternations prototypical to this class (e.g. 104 ↔ 106) but not, for example, in the ones involving PPs (e.g. 103 ↔ 116).

## 47. SUGGEST VERBS

- SCF 16: *John suggested how she could do it*  
 SCF 17: *John suggested how to do it*  
 SCF 24: *John suggested it*  
 SCF 49: *John suggested it to her*  
 SCF 89: *John suggested to her how she could do it*  
 SCF 90: *John suggested to her how to do it*  
 SCF 97: *John suggested to her that she would do it*  
 SCF 98: *John suggested to her that she do it*  
 SCF 101: *John suggested to her what she could do*  
 SCF 103: *John suggested to her what to do*  
 SCF 104: *John suggested that she could do it*  
 SCF 106: *John suggested that she do it*  
 SCF 114: *John suggested what she could do*  
 SCF 116: *John suggested what to do*

**Alternating SCFs:** 16 ↔ 17, 24 ↔ 49, 89 ↔ 16,  
 90 ↔ 17, 97 ↔ 104, 98 ↔ 106, 101 ↔ 114,  
 103 ↔ 116, 104 ↔ 106

Our work resulted in accepting, rejecting, combining and refining the 102 candidate classes and - as a by-product - identifying 5 new classes not included in any of the resources we used. In the end, 57 new verb classes were formed, each associated with 2-45 member verbs. Those Levin or Dorr classes which were examined but found distinctive enough as they stand are not included in this count. However, their possible subclasses are, as well as any of the classes adapted from the resources of Rudanko or Sager. The new classes are listed in table 2, along with example verbs.

## 4 Evaluation

### 4.1 Task-Based Evaluation

We performed an experiment in the context of automatic SCF acquisition to investigate whether the new classes can be used to support an important NLP task. The task is to associate classes to specific verbs along with an estimate of the conditional probability of a SCF given a specific verb. The resulting valency or subcategorization lexicon can be used by a (statistical) parser to recover predicate-argument structure.

Our test data consisted of a total of 35 verbs from 12 new verb classes. The classes were chosen at random, subject to the constraint that their member verbs were frequent enough in corpus data. A minimum of 300 corpus occurrences per verb is required to yield a reliable SCF distribution for a polysemic verb with multiple SCFs (Korhonen, 2002). We took a sample of 20 million words of the British National Corpus (BNC) (Leech, 1992) and extracted all sentences containing an occurrence of one of the test verbs. After the extraction process, we retained

Class	Example Verbs
1. URGE	<i>ask, persuade</i>
2. FORCE	<i>manipulate, pressure</i>
3. ORDER	<i>command, require</i>
4. WANT	<i>need, want</i>
5. TRY	<i>attempt, try</i>
6. WISH	<i>hope, expect</i>
7. ENFORCE	<i>impose, risk</i>
8. ALLOW	<i>allow, permit</i>
9. ADMIT	<i>include, welcome</i>
10. CONSUME	<i>spend, waste</i>
11. PAY	<i>pay, spend</i>
12. FORBID	<i>prohibit, ban</i>
13. REFRAIN	<i>abstain, refrain</i>
14. RELY	<i>bet, count</i>
15. CONVERT	<i>convert, switch</i>
16. SHIFT	<i>resort, return</i>
17. ALLOW	<i>allow, permit</i>
18. HELP	<i>aid, assist</i>
19. COOPERATE	<i>collaborate, work</i>
20. SUCCEED	<i>fail, manage</i>
21. NEGLECT	<i>omit, fail</i>
22. LIMIT	<i>restrict, restrain</i>
23. APPROVE	<i>accept, object</i>
24. ENQUIRE	<i>ask, consult</i>
25. CONFESS	<i>acknowledge, reveal</i>
26. INDICATE	<i>demonstrate, imply</i>
27. DEDICATE	<i>devote, commit</i>
28. FREE	<i>cure, relieve</i>
29. SUSPECT	<i>accuse, condemn</i>
30. WITHDRAW	<i>retreat, retire</i>
31. COPE	<i>handle, deal</i>
32. DISCOVER	<i>hear, learn</i>
33. MIX	<i>pair, mix</i>
34. CORRELATE	<i>coincide, alternate</i>
35. CONSIDER	<i>imagine, remember</i>
36. SEE	<i>notice, feel</i>
37. LOVE	<i>like, hate</i>
38. FOCUS	<i>focus, concentrate</i>
39. CARE	<i>mind, worry</i>
40. DISCUSS	<i>debate, argue</i>
41. BATTLE	<i>fight, communicate</i>
42. SETTLE	<i>agree, contract</i>
43. SHOW	<i>demonstrate, quote</i>
44. ALLOW	<i>allow, permit</i>
45. EXPLAIN	<i>write, read</i>
46. LECTURE	<i>comment, remark</i>
47. SUGGEST	<i>propose, recommend</i>
48. OCCUR	<i>happen, occur</i>
49. MATTER	<i>count, weight</i>
50. AVOID	<i>miss, boycott</i>
51. HESITATE	<i>loiter, hesitate</i>
52. BEGIN	<i>continue, resume</i>
53. STOP	<i>terminate, finish</i>
54. NEGLECT	<i>overlook, neglect</i>
55. CHARGE	<i>commit, charge</i>
56. REACH	<i>arrive, hit</i>
57. ADOPT	<i>assume, adopt</i>

Table 2: New Verb Classes

1000 citations, on average, for each verb.

Our method for SCF acquisition (Korhonen, 2002) involves first using the system of Briscoe and Carroll (1997) to acquire a putative SCF distribution for each test verb from corpus data. This system employs a robust statistical parser (Briscoe and Carroll, 2002) which yields complete though shallow parses from the PoS tagged data. The parse contexts around verbs are passed to a comprehensive SCF classifier, which selects one of the 163 SCFs. The SCF distribution is then smoothed with the back-off distribution corresponding to the semantic class of the predominant sense of a verb. Although many of the test verbs are polysemic, we relied on the knowledge that the majority of English verbs have a single predominating sense in balanced corpus data (Korhonen and Preiss, 2003).

The back-off estimates were obtained by the following method:

- (i) A few individual verbs were chosen from a new verb class whose predominant sense according to the WordNet frequency data belongs to this class,
- (ii) SCF distributions were built for these verbs by manually analysing c. 300 occurrences of each verb in the BNC,
- (iii) the resulting SCF distributions were merged.

An empirically-determined threshold was finally set on the probability estimates from smoothing to reject noisy SCFs caused by errors during the statistical parsing phase.

This method for SCF acquisition is highly sensitive to the accuracy of the lexical-semantic classes. Where a class adequately predicts the syntactic behaviour of the predominant sense of a test verb, significant improvement is seen in SCF acquisition, as accurate back-off estimates help to correct the acquired SCF distribution and deal with sparse data. Incorrect class assignments or choice of classes can, however, degrade performance.

The SCFs were evaluated against manually analysed corpus data. This was obtained by annotating a maximum of 300 occurrences for each test verb in the BNC data. We calculated type precision (the percentage of SCF types that the system proposes which are correct), type recall (the percentage of SCF types in the gold standard that the system proposes) and  $F_1$ -measure<sup>2</sup>. To investigate how well the novel classes help to deal with sparse data, we recorded the total number of SCFs missing in the distributions, i.e. false negatives which did not even occur in the unthresholded distributions and were, therefore, never hypothesized by the parser and classifier. We also compared the similarity between the acquired unthresholded

$${}^2F_1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Measures	Method	
	Baseline	New Classes
Precision (%)	67.1	71.0
Recall (%)	53.9	65.0
$F_1$ -measure (%)	60.0	68.0
RC	0.65	0.74
KL	1.10	0.91
JS	0.90	0.07
CE	2.22	2.10
IS	0.61	0.83
Unseen SCFs	196	115

Table 3: Average results for 35 verbs

and gold standard SCF distributions using several measures of distributional similarity: the Spearman rank correlation (RC), Kullback-Leibler distance (KL), Jensen-Shannon divergence (JS), cross entropy (CE), and intersection (IS)<sup>3</sup>.

Table 3 shows average results for the 35 verbs with the the baseline system and for the system which employs the novel classes. We see that the performance improves when the novel classes are employed, according to all measures used. The method yields 8% absolute improvement in  $F_1$ -measure over the baseline method. The measures of distributional similarity show likewise improved performance. For example, the results with IS indicate that there is a large intersection between the acquired and gold standard SCFs when the method is used, and those with RC demonstrate that the method clearly improves the ranking of SCFs according to the conditional probability distributions of SCFs given each test verb. From the total of 193 gold standard SCFs unseen in the unsmoothed lexicon, only 115 are unseen after using the new classification. This demonstrates the usefulness of the novel classes in helping the system to deal with sparse data.

While these results demonstrate clearly that the new classes can be used to support a critical NLP task, the improvement over the baseline is not as impressive as that reported in (Korhonen, 2002) where Levin’s original classes are employed<sup>4</sup>. While it is possible that the new classes require further adjustment until optimal accuracy can be obtained, it is clear that many of our test verbs (and verbs in our new classes in general) are more polysemic on average and thus more ‘difficult’ than those employed by Korhonen (2002). Our subcategorization acquisition method, based on predominant sense heuristics, is less adequate for these verbs – rather, a method based on word sense disambiguation and the use of multi-

<sup>3</sup>For the details of these measures and their application to this task see Korhonen and Krymolowski (2002).

<sup>4</sup>Korhonen (2002) reports 17.8% absolute improvement in  $F_1$ -measure with the back-off scheme on 45 test verbs.

ple classes should be employed to establish the true upper bound on performance. Korhonen and Preiss (2003) have proposed such a method, but the method is not currently applicable to our test data.

## 4.2 Evaluation of Coverage

Investigating the coverage of the current extended classification over the English verb lexicon is not straightforward because no fully suitable gold standard is available. We conducted a restricted evaluation against the comprehensive semantic classification of WordNet. As WordNet incorporates particularly fine-grained sense distinctions, some of its senses are too idiomatic or marginal for classification at this level of granularity. We aimed to identify and disregard these senses from our investigation.

All the WordNet senses of 110 randomly chosen verbs were manually linked to classes in our extended classification (i.e. to Levin's, Dorr's or our new ones). From the total of 253 senses exemplified in the data, 238 proved suitable (of right granularity) for our evaluation. From these, 21 were left unclassified because no class was found for them in the extended resource. After we evaluated these senses using the method described in section 3, only 7 of them turned out to warrant classes of their own which should be added to the extended classification.

## 5 Discussion

The evaluation reported in the previous section shows that the novel classes can be used to support a NLP task and that the extended classification has good coverage over the English verb lexicon and thus constitutes a resource suitable for large-scale NLP use.

Although the classes resulting from our work can be readily employed for NLP purposes, we plan, in the future, to further integrate them into Levin's taxonomy to yield a maximally useful resource for the research community. While some classes can simply be added to her taxonomy as new classes or subclasses of extant classes (e.g. our 47. SUGGEST VERBS can be added as a subclass to Levin's 37. Verbs of Communication), others will require modifying extant Levin classes. The latter classes are mostly those whose members classify more naturally in terms of their sentential rather than NP and PP complementation (e.g. ones related to Levin's 29. Verbs with Predicative Complements).

This work will require resolving some conflicts between our classification and Levin's. Because lexical-semantic classes are based on partial semantic descriptions manifested in alternations, it is clear that different, equally viable classification schemes can be constructed using the same data and methodology. One can grasp this easily by looking at intersective Levin classes (Dang et al., 1998), created by grouping together subsets of existing classes with overlapping members. Given that there

is strong potential for cross-classification, we will aim to resolve any conflicts by preferring those classes which show the best balance between the accuracy in capturing syntactic-semantic features and the ability to generalize to as many lexical items as possible.

An issue which we did not address in the present work (as we worked on candidate classes), is the granularity of the classification. It is clear that the 'suitable' level of granularity varies from one NLP task to another. For example, tasks which require maximal accuracy from the classification are likely to benefit the most from fine-grained classes (e.g. refined versions of Levin's classes (Green et al., 2001)), while tasks which rely more heavily on the capability of a classification to capture adequate generalizations over a set of lexical items benefit the most from broad classes. Therefore, to provide a general purpose classification suitable for various NLP use, we intend to refine and organize our novel classes into taxonomies which incorporate different degrees of granularity.

Finally, we plan to supplement the extended classification with additional novel information. In the absence of linguistic resources exemplifying further candidate classes we will search for additional novel classes, intersective classes and member verbs using automatic methods, such as clustering (e.g. (Brew and Schulte im Walde, 2002; Korhonen et al., 2003)). For example, clustering sense disambiguated subcategorization data (acquired e.g. from the SemCorpus) should yield suitable (sense specific) data to work with. We will also include in the classification statistical information concerning the relative likelihood of different classes, SCFs and alternations for verbs in corpus data, using e.g. the automatic methods proposed by McCarthy (2001) and Korhonen (2002). Such information can be highly useful for statistical NLP systems utilizing lexical-semantic classes.

## 6 Conclusions

This paper described and evaluated a substantial extension to Levin's widely employed verb classification, which incorporates 57 novel classes and 106 diathesis alternations for verbs not covered comprehensively by Levin. The utility of the novel classes was demonstrated by using them to support automatic subcategorization acquisition. The coverage of the resulting extended classification over the English verb lexicon was shown to be good. Discussion was provided on how the classification could be further refined and extended in the future, and integrated into Levin's extant taxonomy, to yield a single, comprehensive resource.

## Acknowledgements

This work was supported by UK EPSRC project GR/N36462/93: 'Robust Accurate Statistical Parsing (RASP)'.

## References

- B. Boguraev, E. J. Briscoe, J. Carroll, D. Carter, and C. Grover. 1987. The derivation of a grammatically-indexed lexicon from the Longman Dictionary of Contemporary English. In *Proc. of the 25<sup>th</sup> ACL*, pages 193–200, Stanford, CA.
- C. Brew and S. Schulte im Walde. 2002. Spectral clustering for German verbs. In *Conference on Empirical Methods in Natural Language Processing*, Philadelphia, USA.
- E. J. Briscoe and J. Carroll. 1997. Automatic extraction of subcategorization from corpora. In *5<sup>th</sup> ACL Conference on Applied Natural Language Processing*, pages 356–363, Washington DC.
- E. J. Briscoe and J. Carroll. 2002. Robust accurate statistical annotation of general text. In *3<sup>rd</sup> International Conference on Language Resources and Evaluation*, pages 1499–1504, Las Palmas, Gran Canaria.
- E. J. Briscoe. 2000. *Dictionary and System Subcategorisation Code Mappings*. Unpublished manuscript, <http://www.cl.cam.ac.uk/users/alk23/subcat/subcat.html>, University of Cambridge Computer Laboratory.
- H. T. Dang, K. Kipper, M. Palmer, and J. Rosenzweig. 1998. Investigating regular sense extensions based on intersective Levin classes. In *Proc. of COLING/ACL*, pages 293–299, Montreal, Canada.
- B. Dorr. 1997. Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. *Machine Translation*, 12(4):271–325.
- C. Fellbaum. 1999. The organization of verbs and verb concepts in a semantic net. In P. Saint-Dizier, editor, *Predicative Forms in Natural Language and in Lexical Knowledge Bases*, pages 93–110. Kluwer Academic Publishers, Netherlands.
- R. Green, L. Pearl, B. J Dorr, and P. Resnik. 2001. Lexical resource integration across the syntax-semantics interface. In *NAACL Workshop on WordNet and Other Lexical Resources: Applications, Customizations*, CMU, PA.
- R. Grishman, C. Macleod, and A. Meyers. 1994. Complex syntax: building a computational lexicon. In *International Conference on Computational Linguistics*, pages 268–272, Kyoto, Japan.
- R. Jackendoff. 1990. *Semantic Structures*. MIT Press, Cambridge, Massachusetts.
- K. Kipper, H. T. Dang, and M. Palmer. 2000. Class-based construction of a verb lexicon. In *Proc. of the 17th National Conference on Artificial Intelligence*, Austin, TX.
- J. L. Klavans and M. Kan. 1998. Role of verbs in document analysis. In *Proc. of COLING/ACL*, pages 680–686, Montreal, Canada.
- A. Korhonen and Y. Krymolowski. 2002. On the robustness of entropy-based similarity measures in evaluation of subcategorization acquisition systems. In *Proceedings of the 6th Conference on Natural Language Learning*, pages 91–97.
- A. Korhonen and J. Preiss. 2003. Improving subcategorization acquisition using word sense disambiguation. In *Proc. of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan.
- A. Korhonen, Y. Krymolowski, and Z. Marx. 2003. Clustering polysemic subcategorization frame distributions semantically. In *Proc. of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan.
- A. Korhonen. 2002. *Subcategorization Acquisition*. Ph.D. thesis, University of Cambridge, UK.
- B. Levin. 1993. *English Verb Classes and Alternations*. Chicago University Press, Chicago.
- D. McCarthy. 2001. *Lexical Acquisition at the Syntax-Semantics Interface: Diathesis Alternations, Subcategorization Frames and Selectional Preferences*. Ph.D. thesis, University of Sussex, UK.
- P. Merlo and S. Stevenson. 2001. Automatic verb classification based on statistical distributions of argument structure. *Computational Linguistics*, 27(3):373–408.
- G. A. Miller. 1990. WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–312.
- M. Olsen, Bonnie J. Dorr, and Scott C. Thomas. 1997. Toward compact monotonically compositional interlingua using lexical aspect. In *Workshop on Interlinguas in MT*, pages 33–44, San Diego, CA.
- S. Pinker. 1989. *Learnability and Cognition: The Acquisition of Argument Structure*. MIT Press, Cambridge, Massachusetts.
- D. Prescher, S. Riezler, and M. Rooth. 2000. Using a probabilistic class-based lexicon for lexical ambiguity resolution. In *18th International Conference on Computational Linguistics*, pages 649–655, Saarbrücken, Germany.
- J. Rudanko. 1996. *Prepositions and Complement Clauses*. State University of New York Press, Albany.
- N. Sager. 1981. *Natural Language Information Processing*. Addison-Wesley Publishing Company, MA.
- M. Stede. 1998. A generative perspective on verb alternations. *Computational Linguistics*, 24(3):401–430.