Proceedings of the 10th Conference of the European ACL (EACL-03), pp. 163-170.

# A General Feature Space for Automatic Verb Classification

**Eric Joanis** and **Suzanne Stevenson**
Department of Computer Science
University of Toronto
{joanis,suzanne}@cs.toronto.edu

**Abstract**

We develop a general feature space for automatic classification of verbs into lexical semantic classes. Previous work was limited in scope by the need for manual selection of discriminating features, through a linguistic analysis of the target verb classes (Merlo and Stevenson, 2001). We instead analyze the classification structure at a higher level, using the possible defining characteristics of classes as the basis for our feature space. The general feature space achieves reductions in error rates of 42–69%, on a wider range of classes than investigated previously, with comparable performance to feature sets manually selected for the particular classification tasks. Our results show that the approach is generally applicable, and avoids the need for resource-intensive linguistic analysis for each new task.

## 1 Introduction

Wide-coverage language processing systems require large amounts of knowledge about individual words, leading to a lexical acquisition bottleneck. Because verbs play a central role in the syntactic and semantic interpretation of a sentence, much research has focused on automatically learning properties of verbs from text corpora, such as their subcategorization (Brent, 1993; Briscoe and Carroll, 1997), argument roles (Riloff and Schmelzenbach, 1998; Gildea and Jurafsky, 2002), selectional preferences (Resnik, 1996), and lexical semantic classification (Dorr and Jones, 1996; Lapata and Brew, 1999; Schulte im Walde, 2000; Merlo and Stevenson, 2001). Our work aims to extend the applicability of the latter, by developing a general feature space for automatic verb classification.

Specifically, Merlo and Stevenson (2001) showed that verbs could be automatically classified into one of three lexical semantic classes on the basis of five simple statistical features. This work demonstrated the feasibility of verb classification from noisy, easily extractable corpus statistics. However, the approach was limited in scope by the need for manual determination of a discriminating set of features, through a linguistic analysis of the target verb classes. Our work overcomes this limitation by developing a general feature space that avoids the need for individual development of features for specific classes.

The central idea of our approach is as follows. We focus on lexical semantic classes as in Levin (1993) (hereafter Levin), which group together verbs sharing both a common semantics (such as manner of motion or change of state), and a set of syntactic *alternations*. An alternation refers to the alternative mappings of the semantic arguments of a verb to syntactic positions, as in:

1a. *I loaded* <u>*the truck*</u> $\boxed{\textit{with hay}}$.

b. *I loaded* $\boxed{\textit{hay}}$ <u>*onto the truck*</u>.

In Merlo and Stevenson (2001) (hereafter MS01), the differences between the specific semantic arguments and their possible alternations for the three target classes were analyzed to determine a small set of discriminating features. Here, we perform the linguistic analysis at a higher level, by analyzing the range of possible alternations and distinctions among arguments that verbs can exhibit.

As an illustrative example, MS01 determined that an informative statistical feature is the proportion of animate subjects used with a verb, since animacy is correlated with agenthood, and their verb classes differed in whether they have an agentive subject. In our feature space, we generalize this idea to yield a set of features that estimate animacy for each possible syntactic argument slot (not just subject, but direct and indirect object, and object of prepositions). In this way, we capture the observation that, across verb classes in general, semantic arguments in any slot (not just subject) may differ in their proportion of animate entities.

Since we analyze verb class distinctions at this more general level, we need only do the linguistic analysis once, rather than having to do the individual analysis for every set of classes that we want to distinguish. It is worth emphasizing that we do not base our features directly on all the existing Levin classes (cf. Dorr and Jones 1996)—we instead analyze possible alternations for verbs independent of the actual classes. The result is a general classification feature space that in principle is useful for any Levin-type verb classification task.[1]

The generality of our feature space enables experiments on a wider range of classes than has previously been attempted with the approach in MS01 and subsequent extensions (Merlo et al., 2002). We demonstrate the applicability of the feature space to distinctions among 14 classes, including ten new classes in addition to the target classes on which the method was originally developed. To preview our results below, we achieve reductions in error rate ranging from 42% to 69%, averaging more than 50% over all the tasks, demonstrating the potential of the approach for a wide range of classes. Furthermore, the performance compares favourably with sets of features hand-selected for the particular class distinctions, supporting our claim that the general feature space can replace the time-consuming expert linguistic analysis previously needed for each new classification task.

The remainder of the paper describes the analysis underlying the features in our general feature space, the classes chosen for demonstrating its effectiveness, and our experimental procedures and results. We conclude with a discussion of related work, as well as limitations and planned extensions of our approach.

[1]And indeed, it may be useful for other approaches to predicate classification as well, as long as their distinctions have expression in syntactic behaviour (e.g., as in the FrameNet project, `http://www.icsi.berkeley.edu/~framenet/`).

## 2 The Feature Space

Recall that lexical semantic classes as in Levin form a hierarchy of verb groupings with shared meaning and syntax. The constraints on mapping semantic arguments to syntactic positions reflect underlying semantic properties of the verbs (Pinker, 1989; Levin, 1993). Thus alternations, such as those in (1) above, in which the same argument (*hay* or *truck*) is mapped to different positions, are particularly useful in distinguishing verb classes. MS01 further note the importance of the differing semantic (thematic) roles assigned by verbs for distinguishing the classes.

The features in MS01, and those of our general feature space, were designed to tap into the differences between classes both in the alternations they allow, and in the thematic roles assigned. To this end, we generalized the features from MS01, and extended them by considering the full set of alternations from Levin, to capture similar distinctions across a wider range of verb class variability. Specifically, we drew on Part I of Levin, which elaborates the alternations she uses to characterize verb classes, in contrast to Part II, which enumerates the classes themselves.

The feature space is summarized in Table 1, and described in detail in the rest of this section. All frequency counts are normalized by the total number of occurrences of the verb (or, in some cases, a relevant subset).

### 2.1 Features over Syntactic Slots

Syntactic alternations play a central role in defining classes, but are difficult to detect automatically from a corpus (McCarthy, 2000). We have devised several types of features that tap into different aspects of alternation behaviour. One set of features encodes the frequency of the syntactic slots containing verbal arguments in Levin's classes (subject, direct and indirect object, and prepositional phrases), thus approximating the allowable syntac-

| § in Text | Feature Category | #Features |
|---|---|---|
| 2.1 | Syntactic slots | 76 |
| | Slot overlap | 40 |
| | "Empty" words | 4 |
| 2.2 | Passive | 2 |
| | POS of the verb | 6 |
| | Aux, modal, Adv | 13 |
| | Derived forms | 3 |
| 2.3 | Animacy of NPs | 76 |

Table 1: Feature categories with number of features of each type.

tic frames for a verb.[2] For prepositional phrases, we have a separate feature for each of 51 high frequency prepositions, as well as 19 groups of closely related prepositions (e.g., *between*, *in between*, *among*, *amongst*, *amid*, and *amidst* form one group). We also count verb uses where any prepositional phrase occurs.

Using syntactic frame information alone misses critical properties of alternations, since two verbs may occur in the same frames but undergo different mappings of their arguments to the positions. Earlier work has used a measure of overlap over nouns in syntactic slots as a noisy indicator of participation in an alternation (Stevenson and Merlo, 1999; McCarthy, 2000). The idea is that since the same semantic argument may occur in two different slots in a pair of alternating frames, the degree to which those two slots contain the same entities is an indication of the verb's participation in the alternation.

In our feature space, we consider the overlap between each pair of slots that corresponds to an alternation in Part I of Levin (i.e., §1–8). For each alternation in which a semantic argument occurs in one slot in one usage of the verb (such as subject in *The sky cleared*), and in a different slot in the alternant usage (such as object of *from* in *The clouds cleared from the sky*), we add a feature with the measure of overlap in noun (lemma) usage between those two slots (in this case, subject of verb and object of *from*, Levin §2.3.5). When appropriate, we considered prepositions as a group, e.g., when the alternation refers to a general locative or directional specification, rather than to a specific preposition.

Levin also includes a number of alternations in which semantically empty constituents (*it* and *there*) alternate with a contentful subject or object (e.g., *A problem developed/There developed a problem*, Levin §6.1). We added features to count *it* in the subject and direct object positions, and *there* in the subject position. Given our extraction tools, the relevant semantically empty uses of these words are indistinguishable from their contentful usages. However, as with all our features, we assume that the features will be useful even given a certain level of noise, as found by MS01.

### 2.2 Tense, Voice, and Aspect Features

Verb meaning and alternations also interact in interesting ways with voice, tense, and aspect. For example, MS01 used counts of passive and VBN (past participle) usage as redundant indicators of the transitive alternation. Levin notes several alternations that depend on the passive use (§5), so we adopt this as a general feature. Some alternations in Levin indirectly depend on tense of the verb (e.g., the middle voice is usually in the present tense), so we include a set of features that encode the proportion of occurrence of the six POS tags that verbs can take in the Penn tagset: VB, VBP, VBZ, VBG, VBD, and VBN. We also further augment the feature space to count verb uses which include each auxiliary or modal (or any modal), or an adverb, whose use also interacts with voice and aspect. In addition, there are alternations that involve derived forms of the verb (Levin §5.4, §7.1,2), hence we include a set of features that measure the frequency of a verb used as a noun or adjective.

### 2.3 The Animacy Feature

Another feature used by MS01 tapped into properties of the verbal arguments themselves. The animacy feature measured the proportion of subjects that were animate entities (estimated by pronoun use) to capture the distinction between Agent and Theme subjects across their target classes. Here we consider the animacy of each of our syntactic slots. In addition to personal pronouns, we count

[2] Counts of separate slots are less informative than counts of entire syntactic frames, but also easier to extract, and therefore less noisy.

as animate proper NPs labelled as "person" by our chunker (Abney's (1991) SCOL). Other general feature differences across thematic roles are more challenging to detect automatically (e.g., volitionality or independent existence, from Dowty, 1991). We leave further generalization of the idea of distinctive features among thematic roles to future work.

## 3 Experimental Verb Classes

For the experimental investigation of our feature space, we selected a number of verb classes that would be non-trivial test cases for evaluating the method. Both for practical reasons and to make the results of general interest, the classes could neither be too small nor contain mostly infrequent verbs. Pairs or triples of verb classes were selected to form the test pairs/triples for each of a number of separate classification tasks. To illustrate the range of applicability of the feature space, we picked some pairs/triples of classes that were closely related and some that were more dissimilar. Table 2 at the end of this section lists the classes, with their Levin class numbers; their properties are summarized here with examples of each.

Benefactive versus Recipient verbs.

*Mary baked... a cake for Joan/Joan a cake.*

*Mary gave... a cake to Joan/Joan a cake.*

These dative alternation verbs differ in the preposition and the semantic role of its object.

*Admire* versus *Amuse* verbs.

*I admire Jane.* *Jane amuses me.*

These psychological state verbs differ in that the Experiencer argument is the subject of *Admire* verbs, and the object of *Amuse* verbs.

*Run* versus Sound Emission verbs.

*The kids ran in the room./*The room ran with kids.*

*Birds sang in the trees./The trees sang with birds.*

These intransitive activity verbs differ in the prepositional alternations they allow.

*Cheat* versus *Steal* and *Remove* verbs.

*I cheated... Jane of her money/*the money from Jane.*

*I stole... *Jane of her money/the money from Jane.*

These semantically related classes differ in the prepositional alternants they allow.

*Wipe* versus *Steal* and *Remove* verbs.

*Wipe... the dust/the dust from the table/the table.*

*Steal... the money/the money from the bank/*the bank.*

These classes generally allow the same syntactic frames, but differ in the possible semantic role assignment. (Location can be the direct object of *Wipe* verbs but not of *Steal* and *Remove* verbs, as shown.)

*Spray/Load* versus *Fill* versus Other Verbs of Putting (several related Levin classes).

*I loaded... hay on the wagon/the wagon with hay.*

*I filled... *hay on the wagon/the wagon with hay.*

*I put... hay on the wagon/*the wagon with hay.*

These three classes also differ in prepositional alternants. Note, however, that the options for *Spray/Load* verbs overlap with both of the other two types of verbs.

Optionally Intransitive: *Run* versus Change of State versus "Object Drop".

These are the three classes of MS01, which we investigate here for comparison to their results. All are optionally intransitive but assign different semantic roles to their arguments. (Note that the Object Drop verbs are a superset of the Benefactives above.)

The classification tasks we chose vary in difficulty. For many tasks, knowing exactly what PP arguments each verb takes may be sufficient to perform the classification (cf. Dorr and Jones 1996). However, we do not have access to such perfect knowledge, since PP arguments and adjuncts cannot be distinguished with high accuracy. Using our simple extraction tools, for example, the $PP_{for}$ argument in *I admired Jane for her honesty* is not distinguished from the $PP_{for}$ adjunct in *I amused Jane for the money*. Furthermore, PP arguments differ in frequency, so that a highly distinguishing but rarely used alternant will likely not be useful. Differences in PP usage are thus noisy indicators that we expect to be useful but not definitive.

Finally, one of the pairs, *Wipe* versus *Steal* and *Remove* verbs, are not distinguishable by even perfect information about syntactic frames: there is no frame or slot which is allowed by all verbs in one class and no verbs in the other. One might therefore expect this task to be one of the hardest we consider.

| Verb Class | Levin Cl. | # verbs |
|---|---|---|
| Benefactive | 26.1, 26.3 | 35 |
| Recipient | 13.1, 13.3 | 27 |
| *Admire* | 31.2 | 35 |
| *Amuse* | 31.1 | 134 |
| *Run* | 51.3.2 | 79 |
| Sound Emission | 43.2 | 56 |
| *Cheat* | 10.6 | 29 |
| *Steal* and *Remove* | 10.5, 10.1 | 45 |
| *Wipe* | 10.4.1,2 | 35 |
| *Spray/Load* | 9.7 | 36 |
| *Fill* | 9.8 | 63 |
| Other V. of Putting | 9.1–6 | 48 |
| Change of State | 45.1–4 | 169 |
| Object Drop | 26.1,3,7 | 50 |

Table 2: Verb classes (see Section 3), their Levin class numbers, and the number of experimental verbs in each (see Section 4.2).

## 4 Materials and Method

### 4.1 Corpus

We use the British National Corpus (BNC) to estimate all features (Burnard, 2000), a 100M word corpus tagged with the CLAWS POS tagset, which we map to the Penn tagset. It consists of text samples of recent British English ranging over a wide spectrum of domains, including both fiction and non-fiction. Since it is a general corpus, we do not expect any strong domain-specific bias in verb usage.

### 4.2 Verb Selection

We selected our experimental verbs as follows. We started with a list of all the verbs in the given classes from Levin, removing any verb that did not occur at least 100 times in the BNC.[3] Because we assign a single classification to each verb, we also removed any verb: that was deemed excessively polysemous; that belonged to another class under consideration in this study; or for which the class did not correspond to the main sense. (For example, the Recipient-verb sense of *extend*—in the usage "extending a loan"—is far less common than the "stretching" sense.) In on-going work, we are exploring an iterative approach to token- and type-based classification that would address this current limitation of a single class per verb.

Table 2 above shows the number of verbs in each class at the end of this process. Of these verbs, 20 from each class were randomly selected to use as training data. (For two classes, the selection process was not completely random, since some of the verbs were ones used previously in MS01. We used those as training verbs since they did not qualify as unseen test verbs.) Our unseen test verbs then consist of all verbs remaining after the selection of training data, except those that had been used in the previous studies we build on.

### 4.3 Feature Extraction

We use an existing tool to extract the verb group and syntactic slots on which our features are based. The partial (or chunking) parser of Abney (1991), called SCOL, allows us to extract subjects and direct objects with reasonable confidence, and to extract potential prepositional phrases associated with the verb. SCOL does not identify indirect objects, which we also require. We use TGrep2 (Rohde, 2002) to identify potential indirect objects by assuming that an object followed by a noun phrase which SCOL has left unattached is a double-object frame.[4] From these extracted slots, we calculate all the features described in Section 2, yielding a vector of 220 normalized counts for each verb, which forms the input to the machine learning system.

### 4.4 Machine Learning Approach

For all our experiments, we use the decision-tree induction system C5.0, release 1.16 (`www.rulequest.com`), the successor of C4.5 (Quinlan, 1993), as our machine learning software. On our training sets, we use 10-fold cross-validation repeated 50 times to estimate classification accuracy. For these, we report the average accuracy

[3] Due to mistakes in our original counts, twelve verbs with frequency between 88 and 99 were retained, as well as one with frequency 44, *singe*, and one with frequency 19, *idolize*.

[4] This method of indirect object extraction had very low precision (.22 on a random sample), but produced useful features despite the high level of noise.

| Experiment | Base-line Acc | $N$ Test Verbs | All features | | | | Levin-derived subsets | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Train | | Test | | Train | | Test | |
| | | | Acc | SE | Acc | (#) | Acc | SE | Acc | (#) |
| Benefactive/Recipient | 50.0 | 22 | 74.1 | 0.6 | 72.7 | (16) | 79.5 | 0.6 | 86.4 | (19) |
| *Admire*/*Amuse* | 50.0 | 129 | 79.6 | 0.6 | 82.2 | (106) | 80.8 | 0.5 | 90.7 | (117) |
| *Run*/Sound Emission | 50.0 | 79 | 79.9 | 0.7 | 78.5 | (62) | 71.9 | 0.8 | 77.2 | (61) |
| *Cheat*/*Steal–Remove* | 50.0 | 33 | 88.9 | 0.4 | 72.7 | (24) | 83.0 | 0.5 | 72.7 | (24) |
| *Wipe*/*Steal–Remove* | 50.0 | 39 | 79.4 | 0.8 | 84.6 | (33) | 85.8 | 0.6 | 84.6 | (33) |
| *Spray*/*Fill*/Putting | 33.3 | 85 | 78.3 | 0.5 | 65.9 | (56) | 79.2 | 0.5 | 64.7 | (55) |
| Optionally Intrans. | 33.3 | 204 | 68.3 | 0.7 | 72.1 | (147) | 68.7 | 0.5 | 77.0 | (157) |
| 6 Locative Classes | 16.7 | 133 | 69.0 | 0.3 | 56.4 | (75) | 72.8 | 0.4 | 54.9 | (73) |
| 8 Locative Classes | 12.5 | 212 | 69.3 | 0.4 | 59.9 | (127) | 64.7 | 0.3 | 59.0 | (125) |
| All 13 Classes | 7.7 | 507 | 58.6 | 0.3 | 46.4 | (235) | 58.7 | 0.2 | 43.6 | (221) |

Table 3: Experimental Results. Acc is % accuracy; SE is % standard error; # is number of test items correctly classified.

(number of verbs correctly classified over the total number of verbs) and standard error calculated by C5.0. To measure performance on our unseen test data, we train a classifier on the training data, and then measure its accuracy on the test verbs. In all cases, we use the boosting option in C5.0.

## 5 Experimental Results

We performed ten classification tasks, shown in the first column of Table 3. The 2- and 3-way tasks, and their motivation, were described earlier in Section 3. We added three multiway tasks to explore how much we can expect our feature space to scale to multiple class distinctions: The 6-way task involves the *Cheat*, *Steal–Remove*, *Wipe*, *Spray*/*Load*, *Fill*, and "Other Verbs of Putting" classes, all of which undergo similar alternations of locative arguments. To these 6, the 8-way task adds the *Run* and Sound Emission verbs, which also undergo locative alternations. The 13-way task includes all our classes (except Benefactive, which is a subset of Object Drop).

Note that we do not create one classifier which is applied to different test data sets; rather, a separate classifier is defined for each task using only the training data for the verb classes under investigation. In all cases, the training data is balanced, with 20 verbs in each class, so the baseline (chance) performance is $1/k$ for a task discriminating $k$ classes. This baseline is shown in the second column of Table 3, while the third column gives the number of test verbs for each task.

Our first set of experiments uses all our features (as listed in Table 1); the results are shown in columns 4 and 5 of Table 3. In all cases, test performance shows a substantial increase of 22–47 percentage points above the baseline, with a reduction in error rate ranging from 42–69%. Indeed, it is worth noting that what we predicted to be the most difficult task (the *Wipe/Steal–Remove* case) had the best overall performance, of 84.6% (line 5 of the table). In only two of our 2- and 3-way tasks did test performance decrease by more than 5% compared to training performance, indicating good generalizability of the classifiers. The 6- and 8-way tasks had very good performance, and even the 13-way task far exceeded the baseline, although in each case, test performance was substantially less than results on training data. For these complex multiway tasks, it seems, more training verbs would be desirable.

Ideally, we would also like to compare our results to features developed by linguistic experts, such as those in MS01. On our new test verbs in the optionally intransitive task of MS01, our feature space outperformed their hand-crafted features, 72.1% (line 7 of the table) versus 64.2%.[5]

[5]We set out to show that our general feature space could perform comparably to a manually devised one. This result reveals a potential advantage of the general feature set, which

For the other tasks, however, no manually derived feature space exists. We instead compare the general feature space to subsets of our own features (called the Levin-derived subsets) which are hand-selected through an analysis of the classes in Levin.[6] For each class, we have systematically identified the subset of features indicated by the class description given in Levin. For each task, then, the Levin-derived subset is the union of these subsets for all the classes in the task. The results for these feature sets are given in columns 6 and 7 of Table 3. Comparing columns 5 and 7, we see that the general feature space performs as well as or better than the Levin-derived subset on most tasks. For only 3 of the 10 tasks does the subset of features outperform the full feature space by 5% or more.

These results taken together support the hypothesis that our general feature space can be used successfully for automatic verb classification, and can eliminate the need for time-consuming expert analysis for each new task.

## 6 Related Work

Dorr and Jones (1996) showed that perfect knowledge of the allowable syntactic frames for a verb enabled an accuracy of 98% in assigning verbs to Levin classes. We adopt the approach of MS01, which instead approximates such knowledge through statistical corpus analysis, allowing for easier extensibility to new classes. Furthermore, rather than using a class-by-class analysis as in Dorr and Jones (1996), our features are determined through an analysis of the possible alternations for verbs independent of their class assignment, leading to a more general set of features.

Our study generalizes the slot overlap feature of MS01 as an indicator of membership in a verb class, on the assumption that slot overlap is correlated with alternation behaviour. Indeed, McCarthy (2000) confirmed this relation, showing that a slot overlap feature similar to ours could be a useful indicator for a given alternation. McCarthy achieved an increase in performance on her task by using an alternative similarity measure over the selectional preference of a verb for two slots that alternate, indicating that other formulations of slot similarity may be useful to pursue in future work.

Schulte im Walde (2000) and Schulte im Walde and Brew (2002) achieved promising results using unsupervised clustering of verbs in English and German, respectively, according to syntactic frame statistics. Our feature space is more general because it uses a combination of types of features; we have features intended to indicate participation in alternations, and not just subcategorization. While our method used supervised learning, Stevenson and Merlo (1999) found little difference in performance between unsupervised and supervised approaches on the original MS01 task. However, since unsupervised methods are more sensitive to irrelevant features, additional care will be required to determine useful subsets of features from our (larger) general feature space.

Also using unsupervised learning, Oishi and Matsumoto (1997) clustered Japanese verbs automatically into hundreds of semantic classes, which they then combined into a network of 38 classes using linguistic knowledge and semi-automated processing. Their approach, like ours, used a combination of syntactic frame and aspectual features, but a limited set. Our work aims to extend this type of hand-picked feature space to achieve a generality that will limit the need for human expert input in the classification process.

## 7 Conclusion

We achieve very good accuracies across a range of class distinctions, strongly supporting our approach of devising a general feature space for automatic verb classification based on a high-level analysis of verb class distinctions. By basing our features on possible alternations, rather than specific classes, we aim our feature space to be useful not only for those verb classes already developed in Levin, but for any additional classes not yet covered there, that allow for the same alternations. We also believe it will be straightforward to extend the feature space to cover additional alternations

includes a broad range of indicators of syntactic behaviours. These indicators may help to discriminate classes even if they are not conceived of as core distinctions between the classes.

[6]This approach recognizes the important challenge faced by our method in having to automatically select the best features for a task from our large feature space. Admittedly, it does not address the possibility that a linguist may devise a feature for which we have no analogue.

not in the scope of Levin (such as those involving sentential arguments), through the same process of generalizing the existing features to new slots.

Still, our investigation has been limited to English verb classes, and our feature space is limited in being motivated by alternations in English. It thus lacks grammatical features (e.g., Case or other rich morphological properties) used in other languages to mark arguments or indicate the relation between them. Interestingly, Merlo et al. (2002) showed that some of the hand-crafted features of MS01 were useful in Italian for classifying the same verb classes in that language. We think a more general feature space such as ours will have even more potential for crosslinguistic applicability: We have devised a mapping of Levin's analysis of the variation in expression of arguments across classes in English to a general set of features. To the extent that Levin's analysis is grounded in general principles concerning the linking of semantic arguments to their syntactic expression, our feature space is an initial step in capturing variation in the expression of arguments across languages.

## Acknowledgements

We gratefully acknowledge the financial support of NSERC of Canada, Bell University Labs, and the University of Toronto.

## References

Steven Abney. 1991. Parsing by chunks. In Robert Berwick, Steven Abney, and Carol Tenny, editors, *Principle-Based Parsing*. Kluwer Academic Publishers.

Michael Brent. 1993. From grammar to lexicon: Unsupervised learning of lexical syntax. *Computational Linguistics*, 19(3):243–262.

Ted Briscoe and John Carroll. 1997. Automatic extraction of subcategorization from corpora. In *Proceedings of the Fifth ACL Conference on Applied Natural Language Processing (ANLP-97)*, pages 356–363.

Lou Burnard, editor. 2000. *British National Corpus User Reference Guide*.

Bonnie J. Dorr and Doug Jones. 1996. Role of word sense disambiguation in lexical acquisition: Predicting semantics from syntactic cues. pages 322–327.

David R. Dowty. 1991. Thematic proto-roles and argument selection. *Language*, 67(3):547–619.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.

Maria Lapata and Chris Brew. 1999. Using subcategorization to resolve verb class ambiguity. In P. Fung and J. Zhou, editors, *Joint SIGDAT Conference on Empirical Methods in NLP and Very Large Corpora (EMNLP/VLC-99)*, pages 266–274.

Beth Levin. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago UP.

Diana McCarthy. 2000. Using semantic preferences to identify verbal participation in role switching alternations. In *Proceedings of the First Conference of the North American Ch. of the ACL (NAACL-2000)*.

Paola Merlo and Suzanne Stevenson. 2001. Automatic verb classification based on statistical distributions of argument structure. *Computational Linguistics*, 27(3):373–408.

Paola Merlo, Suzanne Stevenson, Vivian Tsang, and Gianluca Allaria. 2002. A multilingual paradigm for automatic verb classification. In *Proceedings of the 40th Annual Meeting of the ACL*, pages 207–214.

Akira Oishi and Yuji Matsumoto. 1997. Detecting the organization of semantic subclasses of japanese verbs. *International Journal of Corpus Linguistics*, 2(1):65–89.

Steven Pinker. 1989. *Learnability and cognition: the acquisition of argument structure*. MIT Press.

J. R. Quinlan. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, CA.

Philip Resnik. 1996. Selectional constraints: an information-theoretic model and its computational realization. *Cognition*, 61(1–2):127–159.

Ellen Riloff and Mark Schmelzenbach. 1998. An empirical approach to conceptual case frame acquisition. In *Proceedings of the Sixth Workshop on Very Large Corpora (WVLC-98)*.

Douglas L. T. Rohde. 2002. TGrep2 user manual version 1.3. Available with the TGrep2 package at `http://tedlab.mit.edu/~dr/Tgrep2/`.

Sabine Schulte im Walde. 2000. Clustering verbs semantically according to their alternation behaviour. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING-2000)*, pages 747–753.

Sabine Schulte im Walde and Chris Brew. 2002. Inducing German semantic verb classes from purely syntactic subcategorisation information. In *Proceedings of the 40th Annual Meeting of the ACL*, pages 223–230.

Suzanne Stevenson and Paola Merlo. 1999. Automatic verb classification using grammatical features. In *Proceedings of the Ninth Conference of the European Chapter of the ACL (EACL-99)*, pages 45–52.