Automatic Lexical Acquisition Based on Statistical Distributions*

Suzanne Stevenson
Department of Computer Science
University of Toronto
6 King's College Road
Toronto, ON Canada M5S 3H5
suzanne@cs.toronto.edu

Paola Merlo
LATL – Department of Linguistics
University of Geneva
2 rue de Candolle
1211 Genève 4 – Suisse
merlo@lettres.unige.ch

Proceedings of COLING-00, pages 815-821

Abstract

We automatically classify verbs into lexical semantic classes, based on distributions of indicators of verb alternations, extracted from a very large annotated corpus. We address a problem which is particularly difficult because the verb classes, although semantically different, show similar surface syntactic behavior. Five grammatical features are sufficient to reduce error rate by more than 50% over chance: we achieve almost 70% accuracy in a task whose baseline performance is 34%, and whose expert-based upper bound we calculated at 86.5%. We conclude that corpus-driven extraction of grammatical features is a promising methodology for fine-grained verb classification.

1 Introduction

Detailed information about verbs is critical to a broad range of NLP and IR tasks, yet its manual determination for large numbers of verbs is difficult and resource intensive. Research on the automatic acquisition of verb-based knowledge has succeded in gleaning syntactic properties of verbs such as subcategorization frames from online resources (Brent, 1993; Briscoe and Carroll, 1997; Dorr, 1997; Manning, 1993). Recently, researchers have investigated statistical corpus-based methods for lexical semantic classification from syntactic properties of verb usage (Aone and McKee, 1996; Lapata and Brew, 1999; Schulte im Walde, 1998; Stevenson and Merlo, 1999; Stevenson et al., 1999; McCarthy, 2000).

Corpus-based approaches to lexical semantic classification in particular have drawn on Levin's hypothesis (Levin, 1993) that verbs can be classified according to the diathesis alternations (alternations in the syntactic expressions of arguments) in which they participate—for example, whether a verb occurs in the dative/prepositional phrase alternation in English. One diagnostic for diathesis alternations is the subcategorization alternatives of a verb. However, some classes exhibit the same subcategorization possibilities but differ in their argument structures, i.e. the content of the thematic roles assigned to the arguments of the verb. This type of situation constitutes a particularly difficult case for corpus-based classification methods.

In this paper, we apply corpus-based lexical acquisition methodology to distinguish classes of verbs which allow the same subcategorizations, but differ in thematic roles. We first assume that one can automatically restrict the choice of classes to those that participate in the relevant subcategorizations (cf. (Lapata and Brew, 1999)). Our proposal is then to use statistics over diathesis alternants as a way to further distinguish those verbs which allow the same subcategorizations, achieving fine-grained classification within that set. Our work focuses on determining the best semantic class for a verb *type*—the set of usages of a verb across a document or corpus—rather than for a single verb *token* in a single local context. In this way, we can exploit the broad behavior of the verb across the corpus to determine its most likely class overall.

We investigate the proposed approach in an in-depth case study of the three major classes of optionally intransitive verbs in English: unergative, unaccusative, and object-drop. More specifically, according to Levin's classification (Levin, 1993), the unergatives are manner of motion verbs, such as *jump* and *march*; the unaccusatives are verbs of change of state, such as *open* and *explode*; the object-drop verbs are unexpressed object alternation verbs, such as *played* and *painted*. These classes all support both transitive and intransitive subcategorizations, but are distinguished by the pattern of thematic role assignments to subject and object position. We automatically classify these verbs on the basis of statistical ap-

* This research was partly sponsored by US NSF grants #9702331 and #9818322, Swiss NSF fellowship 8210-46569, Information Sciences Council of Rutgers University and IRCS, U. of Pennsylvania. This research was conducted while the first author was at Rutgers University.

proximations to syntactic indicators of the underlying argument structures, using numerical features collected from a large syntactically annotated (tagged or parsed) corpus. We apply machine learning techniques to determine whether the frequency distributions of the features, individually or in combination, support automatic classification of the verbs. To preview our results, we demonstrate that combining only five numerical indicators is sufficient to reduce the error rate in this classification task by more than 50% over chance. Specifically, we achieve almost 70% accuracy in a task whose baseline (chance) performance is 34%, and whose expert-based upper bound is calculated at 86.5%. We conclude that a distribution-based method for lexical semantic verb classification is a promising avenue of research.

2 The Argument Structures

Our approach rests on the hypothesis that, even in cases where verb classes cannot be distinguished by subcategorizations, the frequency distributions of syntactic indicators can hold clues to the underlying thematic role differences. We start here then with a description of the subcategorizations and thematic role assignments for each of the three verb classes under investigation.

As optionally intransitive verbs, each of the three classes participates in the transitive/intransitive alternation:

Unergative
(1a) The horse raced past the barn.
(1b) The jockey raced the horse past the barn.
Unaccusative
(2a) The butter melted in the pan.
(2b) The cook melted the butter in the pan.
Object-drop
(3a) The boy washed the hall.
(3b) The boy washed.

Unergatives are intransitive action verbs, as in (1), whose transitive form can be the causative counterpart of the intransitive form. In the causative use, the semantic argument that appears as the subject of the intransitive, as in (1a), surfaces as the object of the transitive, as in (1b) (Hale and Keyser, 1993). Unaccusatives are intransitive change of state verbs, as in (2a); the transitive counterpart for these verbs exhibits the causative alternation, as in (2b). Object-drop verbs, as in (3), have a non-causative transitive/intransitive alternation, in which the object is simply optional.

Classes	Subj of Trans	Obj of Trans	Subj of Intrans
Unergative	Causal Agent	Agent	Agent
Unaccusative	Causal Agent	Theme	Theme
Object-drop	Agent	Theme	Agent

Table 1: Summary of Thematic Alternations.

Each class is distinguished by the content of the thematic roles assigned by the verb. For object-drop verbs, the subject is an Agent and the optional object is a Theme, yielding the thematic assignments (Agent, Theme) and (Agent) for the transitive and intransitive alternants respectively. Unergatives and unaccusatives differ from object-drop verbs in participating in the causative alternation, and also differ from each other in their core thematic argument. In an intransitive unergative, the subject is an Agent, and in an intransitive unaccusative, the subject is a Theme. In the causative transitive form of each, this core semantic argument is expressed as the direct object, with the addition of a Causal Agent (the causer of the action) as subject in both cases. The thematic roles assigned, and their mapping to syntactic position, are summarized in Table 1.

3 The Features for Classification

The key to any automatic classification task is to determine a set of useful features for discriminating the items to be classified. In what follows, we refer to the columns of Table 1 to explain how we expect the thematic distinctions to yield distributional features whose frequencies discriminate among the classes at hand.

Considering column one of Table 1, only unergative and unaccusative verbs assign a Causal Agent to the subject of the transitive. We hypothesize that the causative construction is linguistically more complex than the simple argument optionality of object-drop verbs (Stevenson and Merlo, 1997). We expect then that object-drop verbs will be more frequent in the transitive than the other two classes. Furthermore, the object of an unergative verb receives the Agent role (see the second column of Table 1), a linguistically marked transitive construction (Stevenson and Merlo, 1997). We therefore expect unergatives to be quite rare in the transitive, leading to a three-way distinction in transitive usage among the three classes.

Second, due to the causative alternation of

Table 2: The Features and Their Expected Behavior

Transitivity	Unaccusatives and unergatives have a causative transitive, hence lower transitive use. Furthermore, unergatives have an agentive object, hence very low transitive use.
Passive Voice	Passive implies transitive use, hence correlated with transitive feature.
VBN Tag	Passive implies past participle use (VBN), hence correlated with transitive (and passive).
Causativity	Object-drop verbs do not have a causal agent, hence low "causative" use. Unergatives are rare in the transitive, hence low causative use.
Animacy	Unaccusatives have a Theme subject in the intransitive, hence lower use of animate subjects.

unergatives and unaccusatives, the thematic role of the *subject* of the intransitive is identical to that of the *object* of the transitive, as shown in columns two and three of Table 1. Given the identity of thematic role mapped to subject and object positions, we expect to observe the same noun occurring at times as subject of the verb, and at other times as object of the verb. In contrast, for object-drop verbs, the thematic role of the subject of the intransitive is identical to that of the subject of the transitive, not the object of the transitive. Thus, we expect that it will be less common for the same noun to occur in subject and object position of the same object-drop verb. We hypothesize that this pattern of thematic role assignments will be reflected in differential amount of usage across the classes of the same nouns as subjects and objects for a given verb. Furthermore, since the causative is a transitive use, and the transitive use of unergatives is expected to be rare, this overlap of subjects and objects should primarily distinguish unaccusatives (predicted to have high overlap of subjects and objects) from the other two classes.

Finally, considering columns one and three of Table 1, we note that unergative and object-drop verbs assign an agentive role to their subject in both the transitive and intransitive, while unaccusatives assign an agentive role to their subject only in the transitive. Under the assumption that the intransitive use of unaccusatives is not rare,[1] we then expect that unaccusatives will occur less often overall with an agentive subject than the other two verb classes. On the further assumption that Agents tend to be animate entities more so than Themes, we expect that unaccusatives will occur less frequently with an animate subject compared to unergative and object-drop verbs. Note the importance of our use of frequency distributions: the claim is not that only Agents can be animate, but rather that nouns that receive an Agent role will more often be animate than nouns that receive a Theme role.

The above interactions between thematic roles and the syntactic expressions of arguments thus lead to three features whose distributional properties appear promising for distinguishing the verb classes: transitivity, causativity, and animacy of subject. We also investigate two additional syntactic features, the passive voice and the past participle POS tag (VBN). These features are related to the transitive/intransitive alternation, since a passive use implies a transitive use of the verb, and the use of passive in turn implies the use of the past participle. Our hypothesis is that these five features will exhibit distributional differences in the observed usages of the verbs, which can be used for classification. The features and their expected relevance are summarized in Table 2.

4 Data Collection and Analysis

We chose a set of 20 verbs from each of three classes. The complete list of verbs is reported in Appendix A. Recall that our goal is to achieve a fine-grained classification of verbs that exhibit the same subcategorization frames; thus, the verbs were chosen because they do not generally show massive departures from the intended verb sense (and usage) in the corpus.[2] In order to simplify the counting procedure, we included only the regular ("-ed") simple past/past participle form of the verb, assuming that this would approximate the distribution of the features across all forms of the verb. Finally, as far as we were able given the preceding constraints, we selected verbs that could occur in the transitive and in the passive.

We counted the occurrences of each verb token in a transitive or intransitive use (TRANS), in a

[1] This assumption is based on the linguistic complexity of the causative, and borne out in our corpus analysis.

[2] Though note that there are only 19 unaccusatives because *ripped* was excluded from the analysis as it occurred mostly in a very different use (*ripped off*) in the corpus from the intended change of state usage.

passive or active use (PASS), in a past participle or simple past use (VBN), in a causative or non-causative use (CAUS), and with an animate subject or not (ANIM), as described below. The first three counts (TRANS, PASS, VBN) were performed on the LDC's 65-million word tagged ACL/DCI corpus (Brown, and Wall Street Journal 1987–1989). The last two counts (CAUS and ANIM) were performed on a 29-million word parsed corpus (Wall Street Journal 1988, provided by Michael Collins (Collins, 1997)). The features were counted as follows:

TRANS: The closest noun following a verb was considered a potential object. A verb immediately followed by a potential object was counted as transitive, otherwise as intransitive.

PASS: A token tagged VBD (the tag for simple past) was counted as active. A token tagged VBN (the tag for past participle) was counted as active if the closest preceding auxiliary was *have*, and as passive if the closest preceding auxiliary was *be*.

VBN: The counts for VBN/VBD were based on the POS label in the tagged corpus.

Each of the above counts was normalized over all occurrences of the "-ed" form of the verb, yielding a single relative frequency measure for each verb for that feature.

CAUS: For each verb token, the subject and object (if there was one) were extracted from the parsed corpus, and the proportion of overlap between subject and object nouns across all tokens of a verb was calculated.

ANIM: To approximate animacy without reference to a resource external to the corpus (such as WordNet), we count pronouns (other than *it*) in subject position (cf. (Aone and McKee, 1996)). The assumption is that the words *I*, *we*, *you*, *she*, *he*, and *they* most often refer to animate entities. We automatically extracted all subject/verb tuples, and computed the ratio of occurrences of pronoun subjects to all subjects for each verb.

The aggregate means by class resulting from the counts above are shown in Table 3. The distributions of each feature are indeed roughly as expected according to the description in Section 3. Unergatives show a very low relative frequency of the TRANS feature, followed by unaccusatives, then object-drop verbs. Unaccusative verbs show a high frequency of the CAUS feature and a low frequency of the ANIM feature compared to the other classes. Although expected to be a redundant indicator of transitivity, PASS and VBN do not distinguish between unaccusative and object-drop verbs, indicating that their distributions are sensitive to factors we have not yet investigated.[3]

Table 3: Aggregated Relative Frequency Data for the Five Features. E = unergatives, A = unaccusatives, O = object-drops.

Class	N	MEAN RELATIVE FREQUENCY				
		TRANS	PASS	VBN	CAUS	ANIM
E	20	0.23	0.07	0.21	0.00	0.25
A	19	0.40	0.33	0.65	0.12	0.07
O	20	0.62	0.31	0.65	0.04	0.15

5 Experiments in Classification

The frequency distributions of our features yield a vector for each verb that represents the relative frequency values for the verb on each dimension:

[verb, TRANS, PASS, VBN, CAUS, ANIM, class]

Example: [opened, .69, .09, .21, .16, .36, unacc]

We use the resulting 59 vectors to train an automatic classifier to determine, given a verb that exhibits transitive/intransitive subcategorization frames, which of the three major lexical semantic classes of English optionally intransitive verbs it belongs to. Note that the baseline (chance) performance in this task is 33.9%, since there are 59 vectors and 3 possible classes, with the most common class having 20 verbs.

We used the C5.0 machine learning system (http://www.rulequest.com), a newer version of C4.5 (Quinlan, 1992), which generates decision trees and corresponding rule sets from a training set of known classifications. We found little to no difference in performance between the trees and rule sets, and report only the rule set results. We report here on experiments using a single hold-out training and testing methodology. In this approach, we hold out a single verb vector as the test case, and train the system on the remaining 58 cases. We then test the resulting classifier on the single hold-out case, recording the assigned class for that verb. This is then repeated for each of the 59 verbs. This technique has the benefit of yielding both an overall accuracy rate (when the results are averaged across all 59 trials), as well as providing the data necessary for determining accuracy for each verb class (because we have the classification of each verb when it is the test case). This allows us to evaluate the contribution

[3]These observations have been confirmed by t-tests between feature values for each pair of classes.

Table 4: Percent Accuracy of Verb Classification Task Using Features in Combination. T=TRANS; P=PASS; V=VBN; C=CAUS; An=ANIM. E=unergatives, A=unaccusatives, O=object-drops

	Percent Accuracy by Class			
Features	All	E	A	O
1. T P V C An	69.5	85.0	63.2	60.0
2. P V C An	64.4	80.0	47.4	65.0
3. T V C An	71.2	80.0	73.7	60.0
4. T P C An	61.0	65.0	68.4	50.0
5. T P V An	62.7	70.0	63.2	55.0
6. T P V C	61.0	80.0	42.1	60.0

of individual features with respect to their effect on the performance of individual classes.

We performed experiments on the full set of features, as well as each subset of features with a single feature removed, as reported in Table 4. Consider the first column in the table. The first line shows that the overall accuracy for all five features is 69.5%, a reduction in the error rate of more than 50% above the baseline. The removal of the PASS feature appears to improve performance (row 3 of Table 4). However, it should be noted that this increase in performance results from a single additional verb being classified correctly. The remaining rows show that no feature is superflous or harmful as the removal of any feature has a 5–8% negative effect on performance. Comparable accuracies have been demonstrated using a more thorough cross-validation methodology and using methods that are, in principle, better at taking advantage of correlated features (Stevenson and Merlo, 1999; Stevenson et al., 1999).

The single hold-out protocol provides new data for analysing the performance on individual verbs and classes. The class-by-class accuracies are shown in the remaining columns of Table 4. We can see clearly that, using all five features, the unergatives are classified with much greater accuracy (85%) than the unaccusatives and object-drop verbs (63.2% and 60.0% respectively), as shown in the first row. The remaining rows show that this pattern generally holds for the subsets of features as well, with the exception of line 4.

While future work on our verb classification task will need to focus on determining features that better discriminate unaccusative and object-drop verbs, we can already exclude an explanation of the results based simply on the verbs' or the classes' frequency. Unergatives have the lowest average (log) frequency (1.3), but are the best classified, while unaccusatives and object-drops are comparable (average log frequency = 2). If we group verbs by frequency, the proportion of errors to the total number of verbs remains fairly similar (freq 1: 7 errors/23 verbs; freq. 2: 6 errors/24 verbs; freq. 3: 4 errors/10 verbs). The only verb of frequency 0 is correctly classified, while the only one with log frequency 4 is not . In sum, we do not find that more frequent classes or verbs are more accurately classified.

Importantly, the experiments also enable us to see whether the features indeed contribute to discriminating the classes in the manner predicted in Section 3. The single hold-out results allow us to do this, by comparing the individual class labels assigned using the full set of five features (TRANS, PASS, VBN, CAUS, ANIM) to the class labels assigned using each size four subset of features. This comparison indicates the changes in class labels that we can attribute to the added feature in going from a size four subset to the full set of features. (The individual class labels supporting our analysis below are available from the authors.) We concentrate on the three main features: CAUS, ANIM, TRANS. We find that the behaviour of these features generally does conform to our predictions. We expected that TRANS would help make a three-way distinction among the verb classes. While unergatives are already accurately classified without TRANS, inspection of the change in class labels reveals that the addition of TRANS to the set improves performance on unaccusatives by helping to distinguish them from object-drops. However, in this case, we also observe a loss in precision of unergatives, since some object-drops are now classified as unergatives. Moreover, we expected CAUS and ANIM to be particularly helpful in identifying unaccusatives, and this is also borne out in our analysis of individual labels. We note that the increased accuracy from CAUS is primarily due to better distinguishing unergatives from unaccusatives, and the increased accuracy from ANIM is primarily due to better distinguishing unaccusatives from object-drops. We conclude that the features we have devised are successful in classifing optionally transitive verbs because they capture predicted differences in underlying argument structure. [4]

[4]Matters are more complex with the other two features and we are still interpreting the results. Our prediction

Table 5: Pair-wise Agreement (Calculated by the Kappa Statistics) of Three Experts (E1, E2, E3) Compared to a Gold Standard (Levin) and to the Classifier (Prog). Numbers in parentheses are percentage of verbs on which judges agree.

	Prog	E1	E2	E3
E1	.36 (59)			
E2	.50 (68)	.59 (75)		
E3	.49 (66)	.53 (70)	.66 (77)	
Levin	.54 (69.5)	.56 (71)	.80 (86.5)	.74 (83)

6 Comparison to Experts

In order to evaluate the performance of the algorithm in practice, we need to compare it to the accuracy of classification performed by an expert, which gives a realistic upper bound for the task. In (Merlo and Stevenson, 2000) we report the results of an experiment that measures experts performance and agreement on a classification task very similar to the program we have described here. The results summarised in Table 5 illustrate the performance of the program. On the one hand, the algorithm does not perform at expert level, as indicated by the fact that, for all experts, the lowest agreement score is with the program. On the other hand, the accuracy achieved by the program of 69.5% is only 1.5% less than one of the human experts in comparison to the gold standard. In fact, if we take the best performance achieved by an expert in this task—86.5%—as the maximum achievable accuracy in classification, our algorithm then reduces the error rate over chance by approximately 68%, a very respectable result.

7 Discussion

The work here contributes both to general and technical issues in automatic lexical acquisition.

Firstly, our results confirm the primary role of argument structure in verb classification. Our experimental focus is particularly clear in this regard because we deal with verbs that are "minimal pairs" with respect to argument structure. By classifying verbs that show the same subcategorizations into different classes, we are able to eliminate one of the confounds in classification work created by the fact that subcategorization and argument structure are largely co-variant. We can infer that the accuracy in our classification is due to argument structure information, as subcategorization is the same for all verbs, confirming that the *content* of thematic roles is crucial for classification. Secondly, our results further support the assumption that thematic differences such as these are apparent not only in differences in subcategorization frames, but also in differences in their frequencies. We thus join the many recent results that all seem to converge in supporting the view that the relation between lexical syntax and semantics can be usefully exploited (Aone and McKee, 1996; Dorr, 1997; Dorr and Jones, 1996; Lapata and Brew, 1999; Schulte im Walde, 1998; Siegel, 1998), especially in a statistical framework. Finally, we observe that this information is detectable in a corpus and can be learned automatically. Thus we view corpora, especially if annotated with currently available tools, as useful repositories of implicit grammars.

Technically, our approach extends existing corpus-based learning techniques to a more complex learning problem, in several dimensions. Our statistical approach, which does not require explicit negative examples, extends approaches that encode Levin's alternations directly, as symbolic properties of a verb (Dorr et al., 1995; Dorr and Jones, 1996; Dorr, 1997). We also extend work using surface indicators to approximate underlying properties. (Oishi and Matsumoto, 1997) use case marking particles to approximate grammatical functions, such as subject and object. We improve on this approach by learning argument structure properties, which, unlike grammatical functions, are not marked morphologically. Others have tackled the problem of lexical semantic classification, as we have, but using only subcategorization frequencies as input data (Lapata and Brew, 1999; Schulte im Walde, 1998). By contrast, we explicitly address the definition of features that can tap directly into thematic role differences that are not reflected in subcategorization distinctions. Finally, when learning of thematic role assignment has been the explicit goal, the text has been semantically annotated (Webster and Marcus, 1989), or external semantic re-

was that VBN and PASS would behave similarly to TRANS. In fact, PASS is at best unhelpful in classification. VBN does appear to make the expected three-way distinction. The change in class labels shows that the improvement in performance with VBN results from better distinguishing unergatives from object-drops, and object-drops from unaccusatives. The latter is surprising, since analysis of the data found that the VBN feature values are statistically indistinct for the object-drop and unaccusative classes as a whole.

sources have been consulted (Aone and McKee, 1996). We extend these results by showing that thematic information can be induced from corpus counts.

The experimental results show that our method is powerful, and suited to the classification of lexical items. However, we have not yet addressed the problem of verbs that can have multiple classifications. We think that many cases of ambiguous classification of verb types can be addressed with the notion of intersective sets introduced by (Dang et al., 1998). This is an important concept that proposes that "regular" ambiguity in classification—i.e., sets of verbs that have the same multi-way classifications according to (Levin, 1993)—can be captured with a finer-grained notion of lexical semantic classes. Extending our work to exploit this idea requires only to define the classes appropriately; the basic approach will remain the same. When we turn to consider ambiguity, we must also address the problem that individual instances of verbs may come from different classes. In future research we plan to extend our method to the classification of ambiguous tokens, by experimenting with a function that combines several sources of information: a bias for the verb type (using the cross-corpus statistics we collect), as well as features of the usage of the instance being classified (cf. (Lapata and Brew, 1999; Siegel, 1998)).

References

Chinatsu Aone and Douglas McKee. 1996. Acquiring predicate-argument mapping information in multilingual texts. In Branimir Boguraev and James Pustejovsky, editors, *Corpus Processing for Lexical Acquisition*, pages 191–202. MIT Press.

Michael Brent. 1993. From grammar to lexicon: Unsupervised learning of lexical syntax. *Computational Linguistics*, 19(2):243–262.

Ted Briscoe and John Carroll. 1997. Automatic extraction of subcategorization from corpora. In *Procs of the Fifth ANLP Conference*, pages 356–363.

Michael John Collins. 1997. Three generative, lexicalised models for statistical parsing. In *Procs of ACL'97*, pages 16–23. Madrid, Spain.

Hoa Trang Dang, Karin Kipper, Martha Palmer, and Joseph Rosenzweig. 1998. Investigating regular sense extensions based on intersective Levin classes. In *Procs of COLING-ACL '98*, pages 293–299, Montreal, Canada.

Bonnie Dorr and Doug Jones. 1996. Role of word sense disambiguation in lexical acquisition: Predicting semantics from syntactic cues. In *Proc. of COLING'96*, pages 322–327, Copenhagen, Denmark.

Bonnie Dorr, Joe Garman, and Amy Weinberg. 1995. From syntactic encodings to thematic roles: Building lexical entries for interlingual MT. *Journal of Machine Translation*, 9(3):71–100.

Bonnie Dorr. 1997. Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. *Machine Translation*, 12:1–55.

Ken Hale and Jay Keyser. 1993. On argument structure and the lexical representation of syntactic relations. In K. Hale and J. Keyser, editors, *The View from Building 20*, pages 53–110. MIT Press.

Maria Lapata and Chris Brew. 1999. Using subcategorization to resolve verb class ambiguity. In *Procs of Joint SIGDAT Conference on Empirical Methods in Natural Language*, College Park, MD.

Beth Levin. 1993. *English Verb Classes and Alternations*. University of Chicago Press, Chicago, IL.

Christopher D. Manning. 1993. Automatic acquisition of a large subcategorization dictionary from corpora. In *Procs of ACL'93*, pages 235–242. Ohio State University.

Diana McCarthy. 2000. Using semantic preference to identify verb participation in role switching alternations. In *Procs of NAACL-2000*, Seattle, Washington.

Paola Merlo and Suzanne Stevenson. 2000. Establishing the upper-bound and inter-judge agrement in a verb classification task. In *Procs of LREC-2000*, pages 1659–1664. Athens, Greece.

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

J. Ross Quinlan. 1992. *C4.5 : Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.

Sabine Schulte im Walde. 1998. Automatic semantic classification of verbs according to their alternation behaviour. AIMS Report 4(3), IMS, Universität Stuttgart.

Eric V. Siegel. 1998. *Linguistic Indicators for Language Understanding* Ph.D. thesis, Dept. of Computer Science, Columbia University.

Suzanne Stevenson and Paola Merlo. 1997. Lexical structure and processing complexity. *Language and Cognitive Processes*, 12(1-2):349–399.

Suzanne Stevenson and Paola Merlo. 1999. Verb classification using distributions of grammatical features. In *Procs of EACL'99*. Bergen, Norway.

Suzanne Stevenson, Paola Merlo, Natalia Kariaeva, and Kamin Whitehouse. 1999. Supervised learning of lexical semantic verb classes using frequency distributions. In *Procs of SigLex '99*, College Park, Maryland.

Mort Webster and Mitch Marcus. 1989. Automatic acquisition of the lexical semantics of verbs from sentence frames. In *Procs of ACL'89*, pages 177–184, Vancouver, Canada.

Appendix A

Unergatives: *floated*, *galloped*, *glided*, *hiked*, *hopped*, *hurried*, *jogged*, *jumped*, *leaped*, *marched*, *paraded*, *raced*, *rushed*, *scooted*, *scurried*, *skipped*, *tiptoed*, *trotted*, *vaulted*, *wandered*. Unaccusatives: *boiled*, *changed*, *cleared*, *collapsed*, *cooled*, *cracked*, *dissolved*, *divided*, *exploded*, *flooded*, *folded*, *fractured*, *hardened*, *melted*, *opened*, *simmered*, *solidified*, *stabilized*, *widened*. Object-drops: *borrowed*, *called*, *carved*, *cleaned*, *danced*, *inherited*, *kicked*, *knitted*, *organised*, *packed*, *painted*, *played*, *reaped*, *rented*, *sketched*, *studied*, *swallowed*, *typed*, *washed*, *yelled*.