Proceedings of the 40th ACL, 207-214, 2002

A Multilingual Paradigm for Automatic Verb Classification

Paola Merlo
Linguistics
Univ. of Geneva
merlo@lettres.unige.ch

Suzanne Stevenson **Vivian Tsang**
Computer Science
Univ. of Toronto
{suzanne,vyctsang}@cs.toronto.edu

Gianluca Allaria
Linguistics
Univ. of Geneva
allaria9@etu.unige.ch

Abstract

We demonstrate the benefits of a multilingual approach to automatic lexical semantic verb classification based on statistical analysis of corpora in multiple languages. Our research incorporates two interrelated threads. In one, we exploit the *similarities* in the crosslinguistic classification of verbs, to extend work on English verb classification to a new language (Italian), and to new classes within that language, achieving an accuracy of 86.4% (baseline 33.9%). Our second strand of research exploits the *differences* across languages in the syntactic expression of semantic properties, to show that complementary information about English verbs can be extracted from their translations in a second language (Chinese). The use of multilingual features improves classification performance of the English verbs, achieving an accuracy of 83.5% (baseline 33.3%).

1 Introduction

Automatic development of large scale lexical resources is necessary to support wide coverage and robust NLP tasks. Recent successful approaches in automatic lexical acquisition have extracted knowledge about verbs from large text corpora—including semantic information such as selectional preferences, argument structure, and lexical semantic classification (Aone and McKee, 1996; Lapata and Brew, 1999; McCarthy, 2000; Merlo and Stevenson, 2001; Resnik, 1996; Riloff and Schmelzenbach, 1998; Schulte im Walde, 2000). Thus far, this work has focused primarily on applying methods within a single language. In this paper, we consider the benefits of a multilingual approach to automatic lexical semantic verb classification.

Following the approach proposed by (Merlo and Stevenson, 2001), henceforward MS01, we focus on a semantic classification of verbs based on their argument structure. By argument structure, we mean the thematic roles assigned by a verb, and their mapping to syntactic positions. This provides a similar type of classification to that of (Levin, 1993), although at a coarser level, since it relies solely on general participant role information and not on fine-grained semantics. In the MS01 proposal, the thematic properties of a verb class are analysed in order to develop syntactic indicators that can be extracted from a corpus. The result is a statistical "summary" of the thematic behaviour of a verb, which forms the input to an automatic classification system. We extend this basic methodology within a body of work which uses multilingual corpora to expand both the applicability and the potential performance of the approach.

Taking a multilingual perspective on automatic verb classification has two broad advantages, reflected in our two strands of work within this paradigm. First, we focus on the commonality of the semantic level underlying verbs across languages, with the potential to yield methods that are straightforwardly applicable to multiple languages. For example, on the practical side, we have successfully extended work on verb classification in English directly to the case of Italian, achieving an accuracy of 86.4% on a task whose baseline is 33.9%. On the theoretical side, our work in this vein also opens up the possibility of discovering crosslinguistic lexical regularities and typological generalisations. Indeed, we find that features that were developed for English verbs work very well in Italian, both on the same classes, and on a new class.

The second broad advantage of our multilingual approach is the potential to bring the power of multilingual resources to bear on the problem of verb classification. The basic premise is that a corpus in a second language has information that is not directly available in the corpus for the first language. We hypothesize that, since the underlying semantics of a verb may be expressed differentially in the syntax across different languages, easily extractable features may exist in one language for properties that are not overtly expressed in another language. For example, we have improved the classification of English verbs by augmenting the English feature set with features based on Chinese translations of the verbs, achieving an accuracy of 83.5% (compared to 67.6% using English-only features), on a task with a 33.3% baseline.

These two strands of research exploit crosslinguistic similarities and differences, respectively, in verb behaviour with respect to semantic classes. What unifies the two approaches is the foundation of the method in a feature-based analysis of one of the primitive building blocks of verbal semantics—specifically, thematic roles—and an analysis of the mapping of roles to syntactic positions. Because thematic roles capture crosslinguistically valid properties of arguments, the approach allows, in at least some cases, for thematic-based features to be ported directly across languages. At the same time, crosslinguistic variation in the surface realization of verbal arguments is well-attested, entailing that thematic features from multiple languages have a potential increase in information over monolingual features.

2 The Verb Classes and Features

As in MS01, we define lexical semantic verb classes according to argument structure—that is, the pattern of thematic roles, such as Agent or Theme, that the verbs assign to their arguments, and their mapping to syntactic positions. The use of argument structure provides a semantic level of classification, since it identifies the general roles assigned to participants in the event. Due to the strong correlation between argument structure and subcategorization (Pinker, 1989; Levin, 1993), it is important to show that the method is indeed capturing the semantic level of argument structure and not just subcategorization. Experiments are thus focused on classes which have the same subcategorizations, and are distinguished instead by the content of thematic roles assigned to their arguments.

2.1 Optionally Intransitive Classes

Specifically, we have investigated several English classes whose verbs can appear both transitively and intransitively, but differ in argument structure. While our definition of classes is broader than the fine-grained classification developed by (Levin, 1993), argument structure classes generally correspond to her broader groupings of verbs (such as, e.g., class 45 instead of 45.1 or 45.2). In our case, we look at: manner of motion verbs (Levin's class 51), change of state verbs (Levin's 45), verbs of creation and transformation (Levin's 26), and psychological state verbs (Levin's 31.2).[1]

The classes exhibit the following transitive and intransitive constructions:

Manner of motion:
The lion jumped through the hoop.
The trainer jumped the lion through the hoop.
Change of State:
The butter melted in the pan.
The cook melted the butter in the pan.
Creation/Transformation: (Object Drop)
The contractor built the houses last summer.
The contractor built all summer.[2]
Psychological State:
The rich love their money.
The rich love too.

Table 1 shows that each class is uniquely distinguished by the pattern of thematic roles assigned within these constructions.[3]

	Transitive		Intrans
Classes	Subj	Obj	Subj
MannerOfMotion	CausalAg	Ag	Ag
ChangeOfState	CausalAg	Th	Th
Creation/Trans	Ag	Th	Ag
PsychState	Exp	Stim	Exp

Table 1: Thematic Roles by Class.
Ag=Agent, CausalAg=Causal Agent,
Th=Theme, Exp=Experiencer, Stim=Stimulus

2.2 Preliminary Statistical Features

MS01 investigated the first three of these classes in their monolingual work on English, developing 5 numeric features that encoded summary statistics over the usage of each verb across the

[1] The latter is a specific subclass of Levin's, because this is a case in which her subclasses differ in argument structure.

[2] Note that the progressive, *The contractor was building all summer*, may be more natural in this usage.

[3] Manner of motion and change of state verbs have a causative transitive form (e.g., Levin, 1993), in which the subject argument of the intransitive form becomes the object of the transitive, with the insertion of a Causal Agent as the transitive subject.

Wall Street Journal (WSJ, 65M words).[4] The statistics were shown to approximate the verbs' thematic relations, either directly or indirectly. (Note that all of the features were counts over tagged or parsed text, with no semantic annotation.) The features are: animacy of subject (ANIM), relative frequency of transitive use, calculated in several variants (TRANSitive, PASSive, VBN POS tag), and use in a causative construction (CAUS). We adopted these same features as the starting point for our multilingual work, and refer the interested reader to MS01 for more detail on these initial features.

For each verb, the frequency distributions of the features yield a vector that represents the estimated value of each feature across the entire corpus, such as:

Vector template:
[verb, TRANS, PASS, VBN, CAUS, ANIM, class]
Example:
[open, .69, .09, .21, .16, .36, ChangeOfState]

These vectors are the (supervised) training data for an automatic classifier to determine, given the feature values for a new verb (not from the training set), which of the three classes of verbs it belongs to.

3 Verb Classification in Multiple Languages

The first necessary test of this verb classification approach from the perspective of multilinguality is to verify that the methodology indeed carries over to languages other than English.

3.1 Classes and Features

We selected two of the three classes originally studied for English—change of state and object drop (the more general set of verbs with the same argument structure as the creation/transformation verbs, as noted above)—to show that the same classes could be distinguished within a new language, Italian. We also added the psych verbs, to study whether the method would extend to a new class as well. The psych verbs were chosen as the novel class because they introduce new thematic roles—Experiencer and Stimulus—that had not been previously investigated.

We adopted the three core features used by MS01, ANIM, CAUS, and TRANS, to test whether the features that were developed to distinguish change of state and object drop verbs in English would be useful to make the same class distinction in Italian. We also expected the CAUS and ANIM features to be useful in distinguishing the new class, psych verbs, from change of state verbs. Specifically, in contrast to change of state verbs, psych verbs do not undergo the causative alternation, so the CAUS feature should be lower. Also, the subjects of psych verbs (which are Experiencers) are more likely to be animate than those of change of state verbs, so the ANIM feature should be higher. To distinguish psych verbs from object drop verbs, we note that psych verbs are stative, while object drop verbs are not. We added two features to capture this discriminating aspectual property: PRES, a measure of present tense use, which should be relatively higher for stative verbs, and GERUN, a measure of the gerundive, therefore of progressive use, which is an indicator of non-stativity in both English and Italian (Bertinetto, 1986).

3.2 Materials and Methods

For our experimental set, we chose 20 Italian verbs from each of the three classes; one verb had to be removed later because of initial misclassification, leaving a total of 59 verbs. We were guided in our choice of verbs by the original English verbs studied by MS01 in the change of state and object drop classes, and by the English verbs in the *admire* subclass from (Levin, 1993) for the psych verbs. When Italian translations of these verbs were low frequency (according to (De Mauro, 1993)), we replaced the original translation with a more frequent synonym or antonym.

The feature counts for the 59 verbs were collected from the Italian corpus Parole (`ftp://ftp.ilc.pi.cnr.it/pub/parole`) put at our disposal by the research group of Nicoletta Calzolari, of the CNR of Pisa, who also kindly extracted the initial verb usages for us. This 21-million word corpus provides a representative sample of the Italian language, from newspapers, periodicals and books. Because the corpus is untagged, it was required to extract the features manually; thus, counts were limited to the first 100 occurrences of each verb. (For

[4] The original work on English studied "object drop" verbs, a broader class than creation/transformation verbs, but with the same argument structure. More recent investigations on English have used the more homogeneous creation/transformation class (Joanis and Stevenson, In preparation).

Features Used	Acc (%)	SE (%)
TRANS CAUS ANIM PRES GERUN	85.1	0.3
TRANS CAUS ANIM PRES	85.1	0.3
TRANS CAUS ANIM GERUN	85.4	0.2
TRANS CAUS ANIM	86.4	0.3
TRANS ANIM	86.4	0.3
TRANS CAUS	75.0	0.3
ANIM CAUS	57.4	0.5

Table 2: Classification Accuracy (Acc) and Standard Error (SE) using 10-Fold Crossvalidation (50 repeats)

details of the extraction process, see (Allaria, 2001).)

The TRANS feature is the relative frequency of transitive uses in the sample (i.e., uses followed by direct object). The ANIM feature records the percentage of animate subjects. The CAUS feature records the overlap of subjects and objects, as an indication of participation in the causative alternation. (For more details on the precise calculation of this feature see MS01.) PRES and GERUN record the relative frequency of use of present tense and gerundive in the sample.

3.3 Experimental Results

The result of our data collection is a set of 59 vectors—one set of features per verb—with which we train an automatic classifier for the three verb classes. The baseline for this task is 33.9%. We used the C5.0 machine learning system (http://www.rulequest.com), a newer version of C4.5 (Quinlan, 1992), which is a decision-tree induction algorithm. We ran experiments using both crossvalidation and leave-one-out training and testing methodologies.

Table 2 shows the results of the crossvalidation experiments. First, we note that the classification achieves a very good accuracy, of 86.4%, for a 79% reduction in error rate. These results compare very favorably to the English experiments (which had reached 69.8% accuracy for a somewhat different combination of classes, as described above, with the same baseline accuracy).

Second, we note the unexpected result that all the classification burden appears to be carried by features that were developed for English. This is indicated by the results in Table 2, which shows performance as one feature at a time is removed from the classification input. In fact, only two features—TRANS and ANIM—result in a decrease in classification performance when removed. This is very surprising—not only do the features directly carry over to a new language, but they carry over to a new class of verbs (the psych verbs), which assigns different thematic roles than the roles for which these features were developed.

Table 3 summarizes the results of the leave-one-out experiments, giving F-scores by class, on combinations of TRANS, ANIM, and CAUS. (F is 2PR/(P+R), where P is precision and R is recall.) These results confirm the observations above. CAUS appears to not contribute to learning (line 2). ANIM is crucial for discriminating the change of state and object drop verbs (line 3), while TRANS is critical for all three classes, especially psych verbs (line 4).

Features Used	Psych F	COS F	ObDrp F
1. TRANS CAUS ANIM	.89	.93	.81
2. TRANS ANIM	.89	.93	.81
3. TRANS CAUS	.89	.76	.57
4. ANIM CAUS	0	.70	.62

Table 3: Class-by-Class F-scores using Leave-one-out Methodology. COS=change of state; ObDrp=object drop.

These results show that statistical verb classification, based on argument structure features, has demonstrated usefulness in multiple languages. Moreover, not only the general methodology, but at least some exact features are transferable from one language to another. Furthermore, at least some features generalize to verb class distinctions for which they were not developed. This latter observation is intriguing, as it suggests that thematic roles may be decomposable into more primitive features (cf. (Dowty, 1991)) which our syntactic indicators are tapping into. That is, the ability of features to discriminate new roles may be a function of how much those features reflect underlying commonalities across thematic roles.

4 Verb Classification Using Multilingual Data

Our first strand of work above illustrates one important potential of a multilingual perspective on verb classification, by demonstrating its crosslinguistic applicability. Here we turn to a very different perspective on multilinguality, in which we exploit the power of data from multiple languages (Chinese and English) in training a classifier for verbs in a single language (English). The success of the Italian experiments above rely on a commonality of verb classes and verb behaviour across languages. The work we describe here also relies on an underlying semantic commonality of verbs across the two languages, but uses the crosslinguistic *differences* in how that semantics is expressed to augment the feature set used in classification.

4.1 Classes and Features

We started with the three English classes—manner of motion, change of state, and creation/transformation—and adopted all five of the original features (TRANS, PASS, VBN, CAUS, ANIM) from MS01. We then analysed translations of these types of verbs in Chinese to determine features that potentially discriminate among the three classes. Note that we do not assume that the Chinese translations fall into exactly the same classification structure as the English verbs. Yet, there must be sufficient similarity among the underlying semantics of the verbs across the two languages for the syntactic features in Chinese to be helpful in classifying the original English verbs.

The linguistic analysis of the Chinese verbs led to the determination of the following types of features: Chinese POS tags, passive particles, periphrastic (causative) particles, and various sublexical morphemic properties. The verb tags and particles are overt expressions of semantic information that is not expressed as clearly in English.[5] The verb tags assigned using the POS tagger from the Chinese Knowledge Information Processing Group (CKIP) incorporate both subcategorization and active/stative information (Liu et al., 1995; Tsao, 1996). The particles are overt indicators of the passive construction (an approximate indicator of transitivity, as in English) and of the causative construction (a more reliable version of the English CAUS indicator). The morphemic properties indicate sublexical properties such as the POS of subparts of compound words, and resultative constructions. (More detail on the analysis of the features, and the data collection below, can be found in (Tsang et al., 2002).) We experimented with these features alone, and in combination with the English features, to determine their usefulness in discriminating the English verb classes.

[5]For example, one possible translation for *I cracked an egg* is *Wo (I) jiang (periphrastic particle) dan (egg) da lan (crack)*, in which the periphrastic particle in Chinese explicitly indicates a causative construction.

4.2 Materials and Methods

We chose 20 English verbs from each class, and extracted their features from the British National Corpus (BNC, 100M words), which had been tagged (Brill, 1995) and chunked (Abney, 1996). We collected the Chinese data from a portion of the Mandarin Chinese News Text (MNews, People's Daily and Xinhua newswire sections, 165M characters), from the Linguistic Data Consortium, which was POS-tagged using the CKIP tagger mentioned above. From MNews, we automatically extracted all Chinese compounds with a verb POS-tag, and then selected those that are translations of the 60 English verbs in the appropriate semantic meaning, i.e., manner-of-motion, change-of-state, and creation/transformation. Note that because we are not classifying the Chinese verbs, we can use multiple translations; the average number of translations per English verb is 6.5. The extracted Chinese verbs were manually matched to their English equivalents to form a translation set.

All counts were collected automatically. The Chinese features were calculated as follows. The CKIP feature is the relative frequency of each of the possible verb tags. PASS-PRT and PERI-PRT are the relative frequencies of the passive and causative particles, respectively. A number of features, such as V-V-COMPOUND and V-RES-COMPOUND, are the relative frequencies of various morpheme combinations. Features are calculated together over all the verbs in the translation set of a given English verb. That is, if $C_1, ..., C_i$ are translations of the English verb E_j, then the value of Chinese feature f_k for E_j is the normalized frequency of counts across all occurrences of $C_1, ..., C_i$.

4.3 Experimental Results

The data for each of our machine learning experiments consists of a vector of the relevant features for each verb: we experiment with English data only, Chinese data only, and English and Chinese data combined. We use the resulting vectors as the training data for the C5.0 classification system, as above. Again, we used both 10-fold cross-validation (repeated 50 times) and leave-one-out training methodologies for our experiments. The

Features Used	Acc (%)	SE (%)
All English	61.8	0.4
Best English: ANIM,TRANS	67.6	0.4
All Chinese	78.8	0.3
Best Chinese: CKIP	82.0	0.3
All English & Chinese	81.0	0.5
Best combo: ANIM,TRANS,CKIP	83.5	0.5

Table 4: Classification Accuracy (Acc) and Standard Error (SE) using 10-Fold Crossvalidation (50 repeats)

baseline accuracy is 33.3%.

The key results of the crossvalidation experiments are in Table 4, which shows the performance for all the features in one or both languages, as well as the best combination of features in one or both languages. There are several interesting things to note. First, our main result is that a multilingual set of features outperforms either set of monolingual features on their own. Specifically, the best performance is attained by the combination of the English features ANIM and TRANS, and the Chinese feature CKIP, yielding an accuracy of 83.5% (error rate reduction of 75%). The accuracy rate is significantly different from the next highest accuracy of 82.0% ($p < .01$, t-test with Welch correction, 80df). This is confirmation of our claim that simultaneous use of multilingual data can improve performance in verb classification.

Second, the next best accuracy is attained by a single Chinese feature, the CKIP verb tags. It is remarkable that one feature can so successfully distinguish the three classes; we believe it does so because this set of Chinese POS tags directly captures both syntactic (transitivity) and semantic (active/stative) information; see (Tsang et al., 2002) for more details. This is further evidence that it is extremely helpful to look to multilingual data for increasing the potential of syntactic features in revealing semantic information—in Chinese, a straightforward POS tagger can apparently yield data in the form of a single feature that far outperforms the available English features. (The best performance in MS01 on these same classes was 69.8%.) Thus, this suggests a strategy of choosing multiple languages from which to elicit data, such that the languages are complementary in their syntactic expression of the underlying semantics of verbs.

Finally, it is worth noting that the features that do best monolingually (ANIM, TRANS in English; CKIP in Chinese), are also the best in combination. We have performed numerous experiments, of which the above are only a small set, and though there are some exceptions, this trend generally holds. This is also important, because it allows a selective strategy in feature combination—we can try many features monolingually, and have confidence that those that perform well will also do well in multilingual combinations.

The class-by-class results of the leave-one-out experiments on the combinations of ANIM, TRANS, and CKIP, summarized using F-scores (2PR/(P+R)), are reported in Table 5. The performance confirms that the multilingual features interact to give the best overall performance. While removing CKIP improves performance on manner of motion and change of state verbs, it degrades discrimination of creation/transformation verbs quite a bit (line 2). On the other hand, the removal of ANIM hurts both change of state and creation/transformation verbs but improves manner of motion verbs (line 3). The removal of TRANS somewhat degrades both manner of motion and change of state verbs (line 4). Clearly, the different features are detecting properties of differential benefit to the three classes, and the use of the three together apparently achieves the best balance.

Features Used	MOM F	COS F	C/T F
1. TRANS ANIM CKIP	.93	.90	.92
2. TRANS ANIM	.95	.93	.87
3. TRANS CKIP	.95	.86	.88
4. ANIM CKIP	.90	.88	.92

Table 5: Class-by-Class F-scores using Leave-one-out Methodology. MOM=manner of motion; COS=change of state; C/T=creation/transformation.

5 Related Work

The use of multilingual corpora has been invaluable in several areas of NLP. For example, the underlying commonality of semantics across a parallel corpus has been shown to aid in word sense disambiguation (WSD)—in (Ide, 2000) and (Resnik and Yarowsky, 1999), a parallel corpus was used as a source for lexicalizing some fine-grained English senses. The notion of transferability of information such as syntactic mark-up has also been pursued within parallel corpora (Yarowsky et al., 2001). (Siegel and McKeown, 2000) suggested a potential use of parallel corpora in learning the aspectual classification (i.e., state or event) of English verbs; our results using Chinese (where the verb tags which indirectly indicate such information performed very well) would further encourage such an approach.

However, our multilingual approach does not rest on the use of parallel corpora, and in that sense is perhaps closer to the work of (Dagan and Itai, 1994), which used statistical data from a monolingual corpus to aid in WSD in a different language. We have also taken inspiration from work on Second Language Acquisition, in which evidence of "transfer" of knowledge from a first language when learning a second has been shown to occur in the acquisition of verb class knowledge (e.g., (Helms-Park, 2001; Juffs, 2000)). Finally, both strands of our work have further connections to the machine translation and lexical acquisition work of Dorr and colleagues (e.g., (Dorr, 1993)), which is founded on the notion of underlying semantic commonalities among verbs as the key to crosslinguistic mappings.

6 Conclusions and Future Directions

The multilingual framework discussed here demonstrates the usefulness of verb classification based on the underlying abstract notion of argument structure. This representation captures typological similarities and supports the straightforward extension of an existing classifier to new languages. Moreover, in combination with a method that exploits surface differences between languages, this representation gives rise to significant improvements in performance in the multilingual experiments.

We have established that the general method for automatic verb classification developed for English is indeed directly useful in another language, Italian. Moreover, because they capture the typologically valid notion of thematic relation, the very statistical features that are most useful in English are also most useful in Italian. These results indicate that these languages do share an underlying level at which verbs can be classified similarly. There are also practical benefits from the observation of similarities. One obstacle facing the use of automatic verb classification has been the lack of a definitive classification of verbs in languages other than English (as in Levin, 1993). Our work allows the investigation of the extent to which the same features that distinguish verb classes in English carry over to other languages, yielding a practical method to investigate Levin-type classes crosslinguistically.

One question that naturally arises is whether classes are similar enough across many languages for this method to really work. The results of the experiments that use data from more than one language give an indication. These experiments exploit surface differences, providing varied views of the same data to the learning algorithm, and thus increasing the amount of available information. But the usefulness of this information depends on the features providing different expressions of the same underlying classification—otherwise different types of features would instead give evidence for different spaces of classes and lead to poor results. The biggest practical benefit in using multilingual resources is the considerable increase in available data, not just in size, but in increased richness of information provided to the classifier through the combination of features across languages.

One unexpected result which deserves attention is the cross-class validity of certain features. We think that our indicators are approximate and partial representations of the thematic roles, and therefore pick out only some characteristics of a role and not all. If roles are decomposable, instead of atomic labels, then the defining characteristics can be shared by several roles (cf. (Dowty, 1991)). This would explain why our indicators generalise across classes in a more powerful way than expected. Thus, the apparent ability of certain indicators which were developed for one class, and hence one type of thematic assignment, to become useful indicators for other classes seems to suggest that the inventory of thematic roles that we have explored here should be decomposed into finer-grained primitives. We are currently exploring that approach, along with our on-going investigation of other languages, and additional verb classes.

Acknowledgements

We thank Remo Bindi from CNR Pisa who extracted the Italian data for us, and Eric Joanis from the University of Toronto who extracted the English data. We gratefully acknowledge the generous support of NSERC of Canada, and of the Swiss NSF under grant 11-65328.01.

References

Steven Abney. 1996. Partial parsing via finitestate cascades. In John Carroll, editor, *Proceedings of the Workshop on Robust Parsing (8th ESSLI)*, pages 8–15. Univ. of Sussex, Brighton.

Gianluca Allaria. 2001. Automatic verb classification: Experiments on Italian. Master's thesis, Univ. of Geneva.

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

Pier Marco Bertinetto. 1986. *Tempo, aspetto e azione nel verbo italiano. Il sistema dell'indicativo.* Accademia della Crusca, Firenze.

Eric Brill. 1995. Transformation-based error-driven learning and natural language processing: a case study in part-of-speech tagging. *Computational Linguistics*, 21(4):543–565.

Ido Dagan and Alon Itai. 1994. Word sense disambiguation using a second language monolingual corpus. *Computational Linguistics*, 20(4):563–596.

Tullio De Mauro. 1993. *Il grande dizionario italiano dell'uso.* UTET, Torino.

Bonnie Dorr. 1993. *Machine Translation: A View from the Lexicon.* MIT Press.

David Dowty. 1991. Thematic proto-roles and argument selection. *Language*, 67(3):563–596.

Rena Helms-Park. 2001. Evidence of lexical transfer in learner syntax. *Studies in Second Language Acquisition*, 23(1):71–102.

Nancy Ide. 2000. Cross-lingual sense determination: Can it work? *Computers and the Humanities*, 34:223–234.

Eric Joanis and Suzanne Stevenson. In preparation. A general feature space for automatic verb classification. Manuscript, Univ. of Toronto.

Alan Juffs. 2000. An overview of the second language acquisition of links between verb semantics and morpho-syntax. In J. Archibald, editor, *Second Language Acquisition and Linguistic Theory*, pages 170–179. Blackwell.

Maria Lapata and Chris Brew. 1999. Using subcategorization to resolve verb class ambiguity. In *Proceedings of Joint EMNLP and Workshop on Very Large Corpora*, pages 266–274, College Park, MD.

Beth Levin. 1993. *English Verb Classes and Alternations.* Univ. of Chicago Press.

Shing-Huan Liu, Keh-jiann Chen, Li-ping Chang, and Yeh-Hao Chin. 1995. Automatic part-of-speech tagging for chinese corpora. *Computer Processing of Chinese and Oriental Languages*, 9(1):31–47.

Diana McCarthy. 2000. Using semantic preferences to identify verbal participation in role switching alternations. In *Proceedings of ANLP/NAACL 2000*, pages 256–263, Seattle, WA.

Paola Merlo and Suzanne Stevenson. 2001. Automatic verb classification based on statistical distributions of argument structure. *Computational Linguistics*, 27(3):393–408.

Steven Pinker. 1989. *Learnability and Cognition: The Acquisition of Argument Structure.* MIT Press, Cambridge, MA.

Ross Quinlan. 1992. C4.5 : Programs for machine learning. In *Series in Machine Learning.* Morgan Kaufmann, San Mateo, CA.

Philip Resnik and David Yarowsky. 1999. Distinguishing systems and distinguishing senses: New evaluation methods for word sense disambiguation. *Natural Language Engineering*, 5(2):113–133.

Philip Resnik. 1996. Selectional constraints: an information-theoretic model and its computational realization. *Cognition*, 61(1–2):127–160.

Ellen Riloff and Mark Schmelzenbach. 1998. An empirical approach to conceptual case frame acquisition. In *Proceedings of the Sixth Workshop on Very Large Corpora*, pages 49–56.

Sabine Schulte im Walde. 2000. Clustering verbs semantically according to their alternation behaviour. In *Proceedings of COLING 2000*, pages 747–753, Saarbrücken, Germany.

Eric V. Siegel and Kathleen R. McKeown. 2000. Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights. *Computational Linguistics*, 26(4):595–628.

Vivian Tsang, Suzanne Stevenson, and Paola Merlo. 2002. Crosslinguistic transfer in automatic verb classification. In *Proceedings of COLING 2002*, Taipei, Taiwan. To appear.

Feng-fu Tsao. 1996. On verb classification in Chinese. *Journal of Chinese Linguistics*, 24(1):138–191.

David Yarowsky, Grace Ngai, and Richard Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of HLT 2001.*