Semantically Motivated Subcategorization Acquisition

Anna Korhonen

University of Cambridge, Computer Laboratory

15 JJ Thomson Avenue, Cambridge CB3 0FD, UK

Anna.Korhonen@cl.cam.ac.uk

Abstract

Automatic acquisition of subcategorization lexicons from textual corpora has become increasingly popular. Although this work has met with some success, resulting lexicons indicate a need for greater accuracy. One significant source of error lies in the process of hypothesis selection which is used for removing noise from automatically acquired subcategorization frames (SCFs). In this paper we describe a more accurate semantically-driven approach to hypothesis selection which can be used to improve large-scale SCF acquisition.

1   Introduction

Subcategorization information is vital for successful parsing, however, manual development of large subcategorized lexicons has proved difficult because predicates change behaviour between sublanguages, domains and over time. Additionally, manually developed subcategorization lexicons do not provide the relative frequency of different SCFs for a given predicate, which is essential in a probabilistic approach.

Over the past years acquiring subcategorization dictionaries from textual corpora has become increasingly popular. Different approaches (Brent, 1993; Manning, 1993; Ushioda et al., 1993; Briscoe and Carroll, 1997; Carroll and Rooth, 1998; Sarkar and Zeman, 2000) vary largely according to the methods used and the number of SCFs being extracted. Regardless of this, they perform similarly. They mostly gather information about the syntactic aspects of subcategorization (the type, number and/or relative frequency of SCFs given specific predicates); do not distinguish between various predicate senses, and there is a ceiling on the performance of these systems at around 80% token recall [1].

Subcategorization acquisition proceeds typically in two steps, by (i) generating hypotheses for SCFs and (ii) selecting reliable hypotheses for the final lexicon. Systems vary according to whether raw (Brent, 1993), partially parsed (Manning, 1993; Ushioda et al., 1993) or intermediately parsed (Briscoe and Carroll, 1997; Carroll and Rooth, 1998; Sarkar and Zeman, 2000) corpus data are used as input to the learning process and how cues for hypotheses are defined and identified. Most recent systems use intermediately parsed data as this potentiates more knowledge-based and accurate hypothesis generation. However, as no lexical or semantic information is typically exploited during parsing, the output from hypothesis generation is inevitably noisy and hypothesis selection is needed when aiming for a high accuracy SCF lexicon.

Hypothesis selection is usually done with a hypothesis test, and frequently with a variation of the binomial filter introduced by Brent (1993). Although different modifications to this test have been proposed, both early and recent systems report unreliable performance, particularly

---

[1] Where token recall is the percentage of SCF tokens in a sample of manually analysed text that were correctly acquired by the system.

with low frequency SCFs. According to one account (Briscoe and Carroll, 1997) the majority of errors arise in SCF acquisition because of the statistical filtering process.

Korhonen et al. (2000) investigated reasons for the poor performance of the statistical filters and reported better accuracy with a simple filter which sets a threshold on the relative frequency of SCF entries, selecting only those SCFs which are high in frequency. Korhonen (2000) proposed combining this filter with a method which deals better with sparse data. The method exploits the knowledge that semantically similar verbs show similar subcategorization behaviour. It involves identifying the semantic class of a predicate, using a hypothesis generator to acquire a SCF distribution for the predicate, smoothing this distribution with the back-off estimates of the respective semantic verb class, and setting an empirically defined threshold on the probability estimates from smoothing to filter out unreliable SCFs.

The small scale experiment reported in Korhonen (2000) demonstrates that this semantically-driven approach can significantly improve the accuracy of hypothesis selection, for both high and low frequency SCFs. In this paper, we describe refining this approach further so that it is suitable for larger scale SCF acquisition. Essentially, we propose methods for construction of semantic classes and automatic semantic classification of verbs. We report an experiment which shows that the method can be used to improve large-scale SCF acquisition.

We introduce the baseline SCF acquisition framework in section 2 and describe our extensions to it in section 3. Section 4 gives details of the experimental evaluation and section 5 concludes with directions for future work.

2 Framework for SCF Acquisition

2.1 Hypothesis Generation

We employ for hypothesis generation the SCF acquisition system of Briscoe and Carroll (1997). This system is capable of distinguishing 163 verbal SCFs—a superset of those found in the ANLT (Boguraev et al., 1987) and COMLEX Syntax dic-

tionaries (Grishman et al., 1994)—and returning relative frequencies for each SCF found for a verb.

It works by first tagging, lemmatizing and parsing corpus data using a robust statistical parser (Carroll and Briscoe, 1996) which employs a grammar written in a feature-based unification grammar formalism. This yields complete though intermediate parses. Local syntactic frames including the syntactic categories and head lemmas of constituents are then extracted from parses, from sentence subanalyses which begin/end at the boudaries of predicates.

The resulting patterns are assigned to SCFs on the basis of the feature values of syntactic categories and head lemmas in each pattern. Finally, sets of SCFs are gathered for verbs and putative lexical entries are constructed.

2.2 Hypothesis Selection

The method for hypothesis selection exploits the knowledge that semantically similar verbs are frequently similar also in terms of subcategorization. The motion verbs *fly* and *move*, for example, take similar SCF distributions, which differ essentially from the ones taken e.g. by communication verbs *tell* and *say*.

Although no perfect correspondence exists between the semantic and syntactic properties of verbs, useful generalizations can be made: Levin (1993) has demonstrated that verb *senses* can be divided into semantic classes distinctive in terms of subcategorization. Korhonen (2000) shows that verb *forms* can also be divided into such classes, according to their predominant (i.e. the most frequent) sense.

For instance, the verb form specific SCF distributions for the highly polysemic verbs *fly* and *move* correlate quite well because the predominant senses of these verbs (according to the WordNet (Miller, 1990) frequency data[2]) are similar. They both belong to the Levin "Motion verbs". Good correlation is observed because the distribution of verb senses tends to be zipfian: the predominant sense typically covers most of the total frequency mass (Preiss et al.,

---

[2]In this paper, we refer to the WordNet version 1.6.

2002).

The method for hypothesis selection involves first (manually) identifying, for a single verb, the broad Levin class[3] corresponding to the predominant sense of this verb in WordNet. A SCF distribution is then acquired for the verb from corpus data, using the hypothesis generator described in the previous section. This distribution is smoothed - using linear interpolation (Chen and Goodman, 1996). - with the "back-off" estimates of the Levin class.

Back-off estimates are obtained by (i) choosing 4-5 representative Levin verbs from a class, (ii) building SCF distributions for these verbs by manually analysing c. 300 occurrences of each verb in the BNC corpus (Leech, 1992) and (iii) merging the resulting set of SCF distributions. For example, the back-off estimates for the "Motion verbs" are constructed by merging the SCF distributions for *fly, walk, march* and *travel.*

A simple method is finally used for filtering, which sets an empirically defined threshold on the probability estimates from smoothing.

Essentially, this method involves using a priori knowledge about generalizations of verb semantics to guide subcategorization acquisition. Back-off estimates are used to correct the acquired SCF distribution and deal with sparse data. However, the parameters used in smoothing are obtained by optimising SCF acquisition performance on held-out training data so that most of the smoothed probability is determined by the maximum likelihood estimate (MLE) from the hypothesis generator.

Korhonen (2000) evaluated this method with 60 test verbs from 10 Levin classes. It yielded 88% type precision (the percentage of SCF types that the method proposes which are correct) and 69% type recall (the percentage of SCF types in the gold standard that the method proposes). The method was compared against a baseline method which sets threshold on MLEs of SCFs from the hypothesis generator. F measure[4] was 77.1 for the novel method and 70.1 for the baseline.

This preliminary experiment shows that the proposed method provides an effective way of dealing with low frequency associations and a means of predicting unseen associations in corpus data.

In this paper, we describe how this semantically-driven method for hypothesis selection was refined further and integrated as part of large-scale SCF acquisition.

3 Extensions to the Framework

Applying the method on a large scale requires us to (a) define a comprehensive set of semantic verb classes, (b) to obtain back-off estimates for each class, and (c) to implement a method capable of automatically assigning verbs to semantic classes. We propose methods for (a) and (c) in sections 3.1 and 3.2, respectively. Section 3.3 discusses the work completed on (a), (b) and (c).

3.1 Semantic Classes

Korhonen (2000) classified verbs into Levin classes. These provide us with a good starting point for large-scale SCF acquisition as well. Although not comprehensive, they cover a substantial number of diathesis alternations occurring in English and work on refining and extending this classification is under way (Dang et al., 1998; Dorr, 1997)[5].

As it is important to minimise the cost involved in constructing back-off estimates, we do not adopt the 191 Levin classes as they stand, although this would allow maximal accuracy. We also do not adopt broad Levin classes as they stand, as some proved inaccurate in the preliminary experiment. Rather, a method was devised for determining the specificity of the Levin class(es) required for adequate distinctiveness in terms of subcategorization. The method proceeds in two steps, by examining the:

Step 1: Syntactic similarity between Levin classes. In this, we use Dorr's (1997) source of LDOCE grammatical codes (Procter, 1978) for

---

[3]Korhonen (2000) employs broad Levin classes only (e.g. 51. "Motion verbs"), not subclasses these may divide into (e.g. 51.2 *"Leave* verbs").

[4]$F = \frac{2 \cdot precision \cdot recall}{precision + recall}$

[5]In the work reported in this paper, we concentrate on Levin classes only, leaving the construction of novel verb classes for future work.

Levin classes[6]. Dorr constructed this source by extracting basic syntactic patterns from all the sentences in Levin (1993) and mapping these onto LDOCE codes. To determine syntactic similarity between a set of Levin (sub)classes, we consider the degree of intersection between the LDOCE codes for these classes.

Step 2: Subcategorization similarity between verbs in Levin classes. We choose a few individual verbs whose predominant sense belongs to the Levin (sub)classes under investigation and examine (1) the intersection of SCFs between the verbs, (2) the dissimilarity of SCF distributions for these verbs according to Kullback-Leibler distance (KL) and (3) the similarity in ranking of SCFs in the distributions according to the Spearman rank correlation coefficient (RC)[7].

Step 2 complements Step 1, as the syntactic information included in Levin (1993) is not conclusive and does not provide any information about the relative frequency of SCFs. In addition, it allows us to examine the degree of SCF correlation between all the verb form specific SCF distributions we are actually concerned with.

For example, this method can be used to decide whether the Levin class of "Verbs of Sending and Carrying" is distinctive enough in terms of subcategorization, or whether it should be divided into subclasses, i.e. "*Send*", "*Slide*", "*Bring* and *Take*", "*Carry*" and "*Drive*" verbs. Step 1 determines that the intersection of LDOCE codes between the five subclasses is fairly large. All classes share 2 LDOCE codes, four share 5, and three share 1. Step 2 concludes that the SCF distributions for five individual verbs (one from each subclass: *send, float, bring, carry* and *ferry*) are fairly similar in terms of KL, RC and the intersection of SCFs. The broad class seems thus syntactically coherent enough to provide an adequate basis for back-off estimates[8].

3.2 Automatic Verb Classification

In Korhonen (2000), verbs were manually assigned to semantic classes. For large-scale SCF acquisition, a method is needed for automatic classification of verbs. We propose one which involves assigning verbs to semantic classes via WordNet. Although WordNet's semantic organization does not always go hand in hand with syntactic information, synonymous verbs in WordNet exhibit syntactic behaviour similar to that characterised in the classification system of Levin (Dorr, 1997).

Dorr (1997) has proposed a fully automatic verb classification algorithm which classifies verbs semantically on the basis of their Word-Net synonyms. This algorithm relies solely on lexical resources but is not accurate enough for our purpose. As inaccurate assignments can actually degrade SCF acquisition performance, some allowance for manual intervention is necessary. We therefore propose a semi-automatic algorithm. This assigns entire WordNet synsets to semantic classes[9], making use of Levin's verb index, the LDOCE dictionary and Dorr's LDOCE codes for Levin classes.

The algorithm classifies synsets subhierarchy by subhierarchy, starting from the top level synsets, and going further down in the taxonomy when required. Each synset is classified by first assigning the majority of its member verbs to a semantic class and then choosing the Levin class supported by the highest number of verbs. 'Member verbs' refer here to those which, according to their predominant sense, are members of the synset in question and of its hyponym synsets. The algorithm proceeds as follows:

Step 1: If the majority of member verbs of a given synset S are Levin verbs[10] from the same class, classify S directly.

Step 2: Otherwise, classify more member verbs (according to Step 4a-d) until the majority are classified, and then go back to Step 1.

Step 3: Otherwise, if the classified verbs point to different Levin classes, examine whether S consists of hyponym synsets. If not, assign S to the Levin class supported by the highest number of classified verbs. If yes, go one level down in the hierarchy and classify the hyponym synsets separately, starting again from Step 1.

---

[6]We are indebted to Bonnie Dorr for the use of these codes.

[7]We use SCF distributions obtained via manual analysis of corpus data. See e.g. (Manning and Schütze, 1999) for details of KL and RC.

[8]See Korhonen (2002) for details of this method.

[9]Our objective is to build a static source where Word-Net synsets are associated with different Levin classes. Although static, the source will allow for updating and adding new verbs to WordNet.

[10]'Levin verbs' refer to verbs whose predominant sense belongs to the Levin class in question.

Step 4: If S includes no Levin verbs, proceed as follows to classify the majority of member verbs of S:

(a) Extract the predominant sense of a given verb V from WordNet

(b) Extract the syntactic codes from LDOCE relevant to this sense

(c) Examine whether V could be assigned to a Levin class already associated with the other verbs in the (i) same synset, (i) possible hypernym synset or (iii) possible sister synsets by comparing the LDOCE codes of the sense and Dorr's LDOCE codes of the respective Levin class(es). Given the hypothesised classes, make the final class assignment manually.

(d) If no suitable class is found, re-examine the case after more verbs have been analysed. If the classification remains unsolved, set V aside for later examination.

The above algorithm is for the most part automatic, however, Step 4b and part of Step 4c (the final class assignment) are done manually to ensure accuracy of classification.

For example, the synset no. 00994853 includes 13 member verbs, 4 of which are Levin "Verbs of Sending and Carrying". We need to classify more verbs to determine class assignment. We choose *whisk* and extract its predominant sense from WordNet:

```
whisk -- (move somewhere quickly; "The president
          was whisked away in his limo")
```

In LDOCE the verb has three senses. That corresponding to the predominant WordNet sense is identified as:

```
2. [X9 esp. OFF, AWAY] to remove
   b. by taking suddenly:
      "She whisked the cups away
      / whisked him (off) home"
```

11 Levin classes are already matched with the verbs in the same, hypernym and sister synsets. Those whose syntactic description includes the LDOCE code X9 are:

```
Verbs of putting
Verbs of removing
Verbs of sending and carrying
Verbs of exerting force
Verbs of motion
```

After verifying these options manually, *whisk* is assigned to "Verbs of Sending and Carrying".

3.3 Completed Work

The verb classification algorithm was applied to 3 WordNet verb files (contact, possession and motion verbs). 1581 synsets in these files were assigned to semantic classes and 148 were left unclassified. A small number of synsets (35) from other verb files were classified as well.

From the total of 32 broad Levin classes exemplified among the classified WordNet synsets, 22 of the most frequent were chosen for further work. These were re-grouped into semantic classes by using the method described in section 3.1. This led to the combination of 5 pairs of broad Levin classes and the division of 3 into subclasses. The resulting 20 semantic classes are shown in table 1, labelled by class codes shown in the first column. Back-off estimates for these classes were built using the method described in section 2.2.

	Levin Verbs of
A	9. Putting
B	10. Removing: 10.1-3, 10.5-9
C	10. Removing: 10.4
D	11. Sending and Carrying 12. Exerting Force
E	13. Change of Possession
F	15. Hold and Keep 16. Concealment
G	17. Throwing
H	18. Contact by Impact 19. Poke Verbs
I	20. Contact
J	21. Cutting
K	22. Combining and Attaching: 22.1-4
L	22. Combining and Attaching: 22.5
M	23. Separating and Disassembling: 21.1-3
N	23. Separating and Disassembling: 23.4
O	34. Assessment 35. Searching
P	36. Social Interaction
Q	42. Killing
R	44. Destroy
S	47. Existence: Verbs of Spatial Configuration 50. Assuming Position
T	51. Motion

Table 1: Semantic verb classes

4 Experimental Evaluation

4.1 Classification Algorithm

To evaluate the classification algorithm, we chose 30 synsets from among the 1616 classified. The synsets were chosen at random so that 10 were taken from each of the three WordNet verb files (contact, possession and motion verbs). From the total of 378 verbs in these synsets, we chose for evaluation those 151 which were neither Levin verbs nor classified manually when linking synsets with Levin classes. For these verbs, the semantic classification was compared against manually obtained gold standard classification. The algorithm classified correctly 140 verbs. The accuracy of the class assignment was thus 93%.

This result is encouraging. However, the correspondence between synsets and Levin classes varies largely across WordNet. Further evaluation is therefore needed in the future with a larger set of WordNet files.

4.2 Subcategorization Acquisition

4.2.1 Test Data and Method

To evaluate the revised approach to hypothesis selection, we took a sample of 20M words of BNC. We extracted all sentences containing an occurrence of one of 91 verbs. The verbs were chosen at random, subject to the constraint that they were classified by the algorithm as members of one of the 20 semantic classes constructed (table 1). After the extraction process, we retained c. 1000 citations for each verb.

The sentences containing these verbs were processed by the SCF acquisition system. The hypothesis generator was held constant, the exception being that the data for these experiments were parsed using a probabilistic chart parser (Chitrao and Grishman, 1990). For hypothesis selection, we employed the method of Korhonen (2000) with the extensions described. We also obtained results for the baseline MLE thresholding method without any smoothing.

The results were evaluated against a manual analysis of the corpus data. This was obtained by analysing a maximum of 300 occurrences for each test verb in the BNC corpora. We calculated type precision, type recall and F measure. In addition to the system results, we calculated KL and RC between the acquired unfiltered SCF distributions and the gold standard distributions. We also recorded the total number of unseen SCFs in the acquired unfiltered SCF distributions which occurred in the gold standard distributions. This was to investigate how well the approach deals with sparse data.

4.2.2 Results

Table 2 gives average results for all the 91 test verbs. The semantically-driven method outperforms the baseline on most measures. The improved KL[11] indicates that the method improves the overall accuracy of SCF distributions. The results with RC show that it helps to correct the ranking of SCFs. Precision worsens slightly from the baseline (1.4%), however, recall improves significantly (24%). That both precision and recall are high demonstrates that the method deals well with both highly ranked SCFs and those low in frequency. The baseline method simply ignores low frequency data and therefore yields poor recall. While a total of 114 gold standard SCFs were unseen in the data from the hypothesis generator, only 24 were unseen after smoothing with back-off estimates. This shows further that the method deals effectively with sparse data.

Verb class specific results in table 3 allow us to examine the accuracy of the back-off estimates. The first column shows a semantic verb class and the second indicates the number of verbs tested for the class.

According to the F measure, the semantically-driven method (SEM) outperforms the baseline method (BL) in all 16 verb classes. KL and RC show improvement in 12 classes. Four classes show worse performance: B, E, F, and O. Several reasons were identified for the poor performance (or small improvement) with some verbs/classes: Firstly, the class specific back-off estimates are not accurate enough for highly polysemic verbs.

---

[11] Note that KL ≥ 0, with KL near to 0 denoting strong association, and $-1 \leq$ RC ≤ 1, with RC near to 0 denoting a low degree of association and RC near to -1 and 1 denoting strong association.

Method	KL	RC	System results			Unseen SCFs
			Precision (%)	Recall (%)	F	
Baseline	0.42	0.62	82.2	48.5	61.0	114
Semantic	0.26	0.74	80.8	72.5	76.4	24

Table 2: Average results for 91 verbs

Sem. Class	Verbs Tested	KL		RC		F Measure		Unseen SCFs	
		BL	SEM	BL	SEM	BL	SEM	BL	SEM
A	6	0.39	0.23	0.63	0.83	62.9	83.3	10	2
B	7	0.11	0.21	0.90	0.87	65.7	73.7	8	1
C	3	0.54	0.16	0.33	0.72	54.6	84.0	5	0
D	10	0.63	0.39	0.46	0.62	55.6	74.5	20	2
E	9	0.13	0.16	0.87	0.87	58.6	60.4	17	3
F	5	0.21	0.26	0.78	0.88	50.7	60.2	11	6
G	3	0.52	0.27	0.48	0.88	71.4	89.8	6	1
H	6	0.45	0.27	0.64	0.64	64.9	84.2	4	1
I	3	0.15	0.14	0.59	0.66	57.9	73.7	7	1
J	5	0.40	0.11	0.66	0.84	63.9	84.8	1	1
K	5	0.52	0.45	0.66	0.77	63.3	70.3	3	0
O	4	0.35	0.44	0.59	0.59	54.6	57.7	3	2
P	7	0.55	0.39	0.62	0.63	61.9	65.3	10	3
R	3	0.10	0.00	1.00	1.00	63.2	66.7	0	0
S	4	0.71	0.22	0.37	0.65	66.7	87.9	2	0
T	11	0.59	0.28	0.55	0.71	64.2	89.1	7	1

Table 3: Results for semantic classes

For example, 6 SCFs are reported unseen with class F and these all involve senses not taken by the verbs used for constructing the back-off estimates. Secondly, the back-off estimates are not accurate for verbs whose distribution of senses is not particularly zipfian, i.e. verbs whose predominant sense is not very frequent (e.g. *keep* in class F). Thirdly, the empirically set (verb class specific) filtering thresholds appeared too high/low for some individual verbs. Finally, although most semantic classes proved fine-grained enough, one class appeared too broad: class F, which was obtained by merging two broad Levin classes ("*Hold* and *Keep*" and "Concealment" verbs).

In sum, the method proposed for construction of semantic classes proved fairly accurate. Semantic classes can easily be evaluated in the context of SCF acquisition and refined further where required. Korhonen (2000) predicted that the total number of semantic classes across the whole lexicon is unlikely to exceed 50. This seems still a valid prediction, as promising results were generally obtained assuming a broad notion of a semantic class.

5 Conclusion

We adopted the semantically motivated approach to hypothesis selection proposed by Korhonen (2000) and modified it to be suitable for large-scale SCF acquisition. A method was proposed for construction of semantic classes and a classification algorithm was designed which automatically assigns verbs to semantic classes via WordNet. Experimental evaluation was provided which showed that the classification algorithm is highly accurate and that the revised method for semantically-driven hypothesis selection can be used to succesfully guide, structure and improve the large-scale acquisition of SCFs from corpus data.

Future work will include: investigating reducing manual effort in the classification algorithm and construction of back-off estimates, constructing semantic classes and back-off estimates across the entire lexicon, applying the classification algorithm to the remaining Word-Net synsets and verb files, addressing the prob-

lem of polysemy, and working on improving the accuracy of the filtering method.

References

B. Boguraev, E. J. Briscoe, J. Carroll, D. Carter, and C. Grover. 1987. The derivation of a grammatically-indexed lexicon from the longman dictionary of contemporary english. In *25th Annual Meeting of the Association for Computational Linguistics*, pages 193–200.

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

E. J. Briscoe and J. Carroll. 1997. Automatic extraction of subcategorization from corpora. In *5th ACL Conference on Applied Natural Language Processing*, pages 356–363.

J. Carroll and E. J. Briscoe. 1996. Apportioning development effort in a probabilistic lr parsing system through evaluation. In *1st ACL/SIGDAT Conference on Empirical Methods in Natural Language Processing*, pages 92–100.

G. Carroll and M. Rooth. 1998. Valence induction with a head-lexicalized pcfg. In *3rd Conference on Empirical Methods in Natural Language Processing*.

Stanley F. Chen and Joshua Goodman. 1996. An empirical study of smoothing techniques for language modeling. In *ACL-96*, pages 310–318. ACL.

M. Chitrao and R. Grishman. 1990. Statistical parsing of messages. In *Darpa Speech and Natural Language Workshop*, pages 263–266.

H. T. Dang, K. Kipper, M. Palmer, and J. Rosensweig. 1998. Investigating regular sense extensions based on intersective levin classes. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, pages 293–299.

B. Dorr. 1997. Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. *Machine Translation*, 12(4):271–325.

R. Grishman, C. Macleod, and A. Meyers. 1994. Comlex syntax: building a computational lexicon. In *International Conference on Computational Linguistics*, pages 268–272.

A. Korhonen, G. Gorrell, and D. McCarthy. 2000. Statistical filtering and subcategorization frame acquisition. In *Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, pages 199–205.

A. Korhonen. 2000. Using semantically motivated estimates to help subcategorization acquisition. In *Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, pages 216–223.

A. Korhonen. 2002. *Subcategorization Acquisition*. Ph.D. thesis, University of Cambridge, UK.

G. Leech. 1992. 100 million words of english: the british national corpus. *Language Research*, 28(1):1–13.

B. Levin. 1993. *English Verb Classes and Alternations*. Chicago University Press, Chicago.

C. Manning and H. Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, Massachusetts.

C. Manning. 1993. Automatic acquisition of a large subcategorization dictionary from corpora. In *31st Annual Meeting of the Association for Computational Linguistics*, pages 235–242.

G. A. Miller. 1990. Wordnet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–312.

J. Preiss, A. Korhonen, and E. J. Briscoe. 2002. Subcategorization acquisition as an evaluation method for wsd. In *Language Resources and Evaluation Conference*. To appear.

P. Procter. 1978. *Longman Dictionary of Contemporary English*. Longman, England.

A. Sarkar and D. Zeman. 2000. Automatic extraction of subcategorization frames for czech. In *19th International Conference on Computational Linguistics*, pages 691–697.

A. Ushioda, D. Evans, T. Gibson, and A. Waibel. 1993. The automatic acquisition of frequencies of verb subcategorization frames from tagged corpora. In B. Boguraev and J. Pustejovsky, editors, *SIGLEX ACL Workshop on the Acquisition of Lexical Knowledge from Text*, pages 95–106. Columbus, Ohio.