A System for Large-Scale Acquisition of Verbal, Nominal and Adjectival Subcategorization Frames from Corpora

Judita Preiss, Ted Briscoe, and Anna Korhonen
Computer Laboratory
University of Cambridge
15 JJ Thomson Avenue
Cambridge CB3 0FD, UK
Judita.Preiss, Ted.Briscoe, Anna.Korhonen@cl.cam.ac.uk

Abstract

This paper describes the first system for large-scale acquisition of subcategorization frames (SCFs) from English corpus data which can be used to acquire comprehensive lexicons for verbs, nouns and adjectives. The system incorporates an extensive rule-based classifier which identifies 168 verbal, 37 adjectival and 31 nominal frames from grammatical relations (GRs) output by a robust parser. The system achieves state-of-the-art performance on all three sets.

1 Introduction

Research into automatic acquisition of lexical information from large repositories of unannotated text (such as the web, corpora of published text, etc.) is starting to produce large scale lexical resources which include frequency and usage information tuned to genres and sublanguages. Such resources are critical for natural language processing (NLP), both for enhancing the performance of state-of-art statistical systems and for improving the portability of these systems between domains.

One type of lexical information with particular importance for NLP is subcategorization. Access to an accurate and comprehensive subcategorization lexicon is vital for the development of successful parsing technology (e.g. (Carroll et al., 1998), important for many NLP tasks (e.g. automatic verb classification (Schulte im Walde and Brew, 2002)) and useful for any application which can benefit from information about predicate-argument structure (e.g. Information Extraction (IE) ((Surdeanu et al., 2003)).

The first systems capable of automatically learning a small number of verbal subcategorization frames (SCFs) from unannotated English corpora emerged over a decade ago (Brent, 1991; Manning, 1993). Subsequent research has yielded systems for English (Carroll and Rooth, 1998; Briscoe and Carroll, 1997; Korhonen, 2002) capable of detecting comprehensive sets of SCFs with promising accuracy and demonstrated success in application tasks (e.g. (Carroll et al., 1998; Korhonen et al., 2003)). Recently, a large publicly available subcategorization lexicon was produced using such technology which contains frame and frequency information for over 6,300 English verbs – the VALEX lexicon (Korhonen et al., 2006).

While there has been considerable work in the area, most of it has focussed on verbs. Although verbs are the richest words in terms of subcategorization and although verb SCF distribution data is likely to offer the greatest boost in parser performance, accurate and comprehensive knowledge of the many noun and adjective SCFs in English could improve the accuracy of parsing at several levels (from tagging to syntactic and semantic analysis).

Furthermore the selection of the correct analysis from the set returned by a parser which does not initially utilize fine-grained lexico-syntactic information can depend on the *interaction* of conditional probabilities of lemmas of different classes occur-

Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pages 912–919,
Prague, Czech Republic, June 2007.

ring with specific SCFs. For example, a) and b) below indicate the most plausible analyses in which the sentential complement attaches to the noun and verb respectively

a) Kim (VP believes (NP the evidence (Scomp that Sandy was present)))

b) Kim (VP persuaded (NP the judge) (Scomp that Sandy was present))

However, both a) and b) consist of an identical sequence of coarse-grained lexical syntactic categories, so correctly ranking them requires learning that $P(NP \mid believe).P(Scomp \mid evidence) > P(NP\&Scomp \mid believe).P(None \mid evidence)$ and $P(NP \mid persuade).P(Scomp \mid judge) < P(NP\&Scomp \mid persuade).P(None \mid judge)$. If we acquired frames and frame frequencies for all open-class predicates taking SCFs using a single system applied to similar data, we would have a better chance of modeling such interactions accurately.

In this paper we present the first system for large-scale acquisition of SCFs from English corpus data which can be used to acquire comprehensive lexicons for verbs, nouns and adjectives. The classifier incorporates 168 verbal, 37 adjectival and 31 nominal SCF distinctions. An improved acquisition technique is used which expands on the ideas Yallop et al. (2005) recently explored for a small experiment on adjectival SCF acquisition. It involves identifying SCFs on the basis of grammatical relations (GRs) in the output of the RASP (Robust Accurate Statistical Parsing) system (Briscoe et al., 2006).

As detailed later, the system performs better with verbs than previous comparable state-of-art systems, achieving 68.9 F-measure in detecting SCF types. It achieves similarly good performance with nouns and adjectives (62.2 and 71.9 F-measure, respectively).

Additionally, we have developed a tool for linguistic annotation of SCFs in corpus data aimed at alleviating the process of obtaining training and test data for subcategorization acquisition. The tool incorporates an intuitive interface with the ability to significantly reduce the number of frames presented to the user for each sentence.

We introduce the new system for SCF acquisition in section 2. Details of the experimental evaluation are supplied in section 3. Section 4 provides discussion of our results and future work, and section 5 concludes.

2 Description of the System

A common strategy in existing large-scale SCF acquisition systems (e.g. (Briscoe and Carroll, 1997)) is to extract SCFs from parse trees, introducing an unnecessary dependence on the details of a particular parser. In our approach SCFs are extracted from GRs — representations of head-dependent relations which are more parser/grammar independent but at the appropriate level of abstraction for extraction of SCFs.

A similar approach was recently motivated and explored by Yallop et al. (2005). A decision-tree classifier was developed for 30 adjectival SCF types which tests for the presence of GRs in the GR output of the RASP (Robust Accurate Statistical Parsing) system (Briscoe and Carroll, 2002). The results reported with 9 test adjectives were promising (68.9 F-measure in detecting SCF types).

Our acquisition process consists of four main steps: 1) extracting GRs from corpus data, 2) feeding the GR sets as input to a rule-based classifier which incrementally matches them with the corresponding SCFs, 3) building lexical entries from the classified data, and 4) filtering those entries to obtain a more accurate lexicon. The details of these steps are provided in the subsequent sections.

2.1 Obtaining Grammatical Relations

We obtain the GRs using the recent, second release of the RASP toolkit (Briscoe et al., 2006). RASP is a modular statistical parsing system which includes a tokenizer, tagger, lemmatizer, and a wide-coverage unification-based tag-sequence parser. We use the standard scripts supplied with RASP to output the set of GRs for the most probable analysis returned by the parser or, in the case of parse failures, the GRs for the most likely sequence of subanalyses. The GRs are organized as a subsumption hierarchy as shown in Figure 1.

The dependency relationships which the GRs embody correspond closely to the head-complement structure which subcategorization acquisition attempts to recover, which makes GRs ideal input to the SCF classifier. Consider the arguments of *easy*

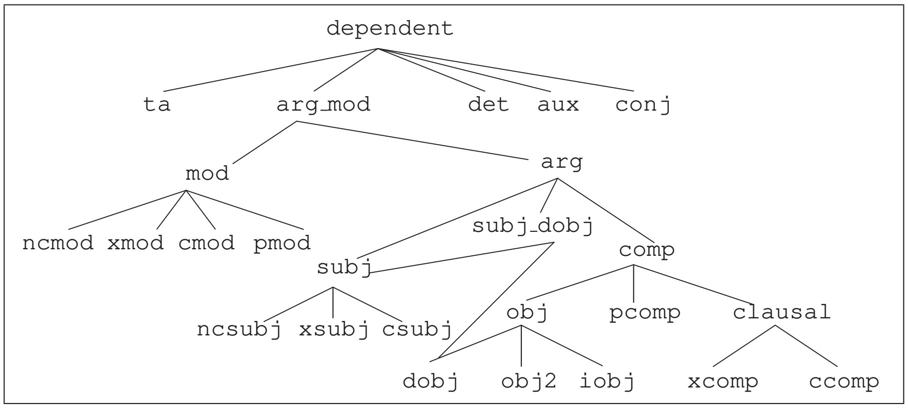

Figure 1: The GR hierarchy used by RASP

$$\begin{bmatrix} \text{SUBJECT} & \text{NP}_{\boxed{1}}, \\ \text{ADJ-COMPS} & \left\langle {}_{\text{PP}}\begin{bmatrix}\text{PVAL} & \text{for}\\ \text{NP} & \boxed{3}\end{bmatrix}, {}_{\text{VP}}\begin{bmatrix}\text{MOOD} & \text{to-infinitive}\\ \text{SUBJECT} & \boxed{3} \\ \text{OMISSION} & \boxed{1}\end{bmatrix} \right\rangle \end{bmatrix}$$

Figure 2: Feature structure for SCF `adj-obj-for-to-inf`

```
(|These:1_DD2| |example+s:2_NN2| |of:3_IO|
 |animal:4_JJ| |senses:5_NN2| |be+:6_VBR|
 |relatively:7_RR| |easy:8_JJ| |for:9_IF|
 |we+:10_PPIO2| |to:11_TO| |comprehend:12_VV0|)
...
 xcomp(_                 be+[6]        easy:[8])
 xcomp(to[11]            be+[6]        comprehend:[12])
ncsubj(be+[6]            example+s[2]  _)
 ncmod(for[9]            easy[8]       we+[10])
ncsubj(comprehend[12]    we+[10], _)
...
```

Figure 3: GRs from RASP for `adj-obj-for-to-inf`

xcomp	_	?Y : pos=vb,val=be	?X : pos=adj
xcomp	?S : val=to	?Y : pos=vb,val=be	?W : pos=VV0
ncsubj	?Y : pos=vb,val=be	?Z : pos=noun	
ncmod	?T : val=for	?X : pos=adj	?Y: pos=pron
ncsubj	?W : pos=VV0	?V : pos=pron	

Figure 4: Pattern for frame `adj-obj-for-to-inf`

in the sentence: *These examples of animal senses are relatively easy for us to comprehend as they are not too far removed from our own experience.* According to the COMLEX classification, this is an example of the frame `adj-obj-for-to-inf`, shown in Figure 2, (using AVM notation in place of COMLEX s-expressions). Part of the output of RASP for this sentence is shown in Figure 3.

Each instantiated GR in Figure 3 corresponds to one or more parts of the feature structure in Figure 2. `xcomp(_ be[6] easy[8])` establishes `be[6]` as the head of the VP in which `easy[8]` occurs as a complement. The first (PP)-complement is *for us*, as indicated by `ncmod(for[9] easy[8] we+[10])`, with *for* as PFORM and `we+` (*us*) as NP. The second complement is represented by `xcomp(to[11] be+[6] comprehend[12])`: a *to-infinitive* VP. The NP headed by *examples* is marked as the subject of the frame by `ncsubj(be[6] examples[2])`, and `ncsubj(comprehend[12] we+[10])` corresponds to the coindexation marked by $\boxed{3}$: the subject of the VP is the NP of the PP. The only part of the feature structure which is not represented by the GRs is coindexation between the omitted direct object $\boxed{1}$ of the VP-complement and the subject of the whole clause.

2.2 SCF Classifier

SCF Frames

The SCFs recognized by the classifier were obtained by manually merging the frames exemplified in the COMLEX Syntax (Grishman et al., 1994), ANLT (Boguraev et al., 1987) and/or NOMLEX (Macleod et al., 1997) dictionaries and including additional frames found by manual inspection of unclassifiable examples during development of the classifier. These consisted of e.g. some occurrences of phrasal verbs with complex complementation and with flexible ordering of the preposition/particle, some non-passivizable words with a surface direct object, and some rarer combinations of governed preposition and complementizer combinations.

The frames were created so that they abstract over specific lexically-governed particles and prepositions and specific predicate selectional preferences but include some derived semi-predictable bounded dependency constructions.

Classifier

The classifier operates by attempting to match the set of GRs associated with each sentence against one or more rules which express the possible mappings from GRs to SCFs. The rules were manually developed by examining a set of development sentences to determine which relations were actually emitted by the parser for each SCF.

In our rule representation, a GR pattern is a set of partially instantiated GRs with variables in place of heads and dependents, augmented with constraints that restrict the possible instantiations of the variables. A match is successful if the set of GRs for a sentence can be unified with any rule. Unification of sentence GRs and a rule GR pattern occurs when there is a one-to-one correspondence between sentence elements and rule elements that includes a consistent mapping from variables to values.

A sample pattern for matching `adj-obj-for-to-inf` can be seen in Figure 4. Each element matches either an empty GR slot (_), a variable with possible constraints on part of speech (pos) and word value (val), or an already instantiated variable. Unlike in Yallop's work (Yallop et al., 2005), our rules are declarative rather than procedural and these rules, written independently of the acquisition system, are expanded by the system in a number of ways prior to execution. For example, the verb rules which contain an `ncsubj` relation will not contain one inside an embedded clause. For verbs, the basic rule set contains 248 rules but automatic expansion gives rise to 1088 classifier rules for verbs.

Numerous approaches were investigated to allow an efficient execution of the system: for example, for each target word in a sentence, we initially find the number of ARGument GRs (see Figure 1) containing it in head position, as the word must appear in exactly the same set in a matching rule. This allows us to discard all patterns which specify a different number of GRs: for example, for verbs each group only contains an average of 109 patterns.

For a further increase in speed, both the sentence GRs and the GRs within the patterns are ordered (according to frequency) and matching is performed using a backing off strategy allowing us to exploit the relatively low number of possible GRs (compared to the number of possible rules). The system executes on 3500 sentences in approx. 1.5 seconds of real time on a machine with a 3.2 GHz Intel Xenon processor and 4GB of RAM.

Lexicon Creation and Filtering

Lexical entries are constructed for each word and SCF combination found in the corpus data. Each lexical entry includes the raw and relative frequency of the SCF with the word in question, and includes various additional information e.g. about the syntax of detected arguments and the argument heads in different argument positions[1].

Finally the entries are filtered to obtain a more accurate lexicon. A way to maximise the accuracy of the lexicon would be to smooth (correct) the acquired SCF distributions with back-off estimates based on lexical-semantic classes of verbs (Korhonen, 2002) (see section 4) before filtering them. However, in this first experiment with the new system we filtered the entries directly so that we could evaluate the performance of the new classifier without any additional modules. For the same reason, the filtering was done by using a very simple method: by setting empirically determined thresholds on the relative frequencies of SCFs.

3 Experimental Evaluation

3.1 Data

In order to test the accuracy of our system, we selected a set of 183 verbs, 30 nouns and 30 adjectives for experimentation. The words were selected at random, subject to the constraint that they exhibited multiple complementation patterns and had a sufficient number of corpus occurrences ($>$ 150) for experimentation. We took the 100M-word British National Corpus (BNC) (Burnard, 1995), and extracted all sentences containing an occurrence of one of the test words. The sentences were processed using the SCF acquisition system described in the previous section. The citations from which entries were derived totaled approximately 744K for verbs and 219K for nouns and adjectives, respectively.

[1]The lexical entries are similar to those in the VALEX lexicon. See (Korhonen et al., 2006) for a sample entry.

3.2 Gold Standard

Our gold standard was based on a manual analysis of some of the test corpus data, supplemented with additional frames from the ANLT, COMLEX, and/or NOMLEX dictionaries. The gold standard for verbs was available, but it was extended to include additional SCFs missing from the old system. For nouns and adjectives the gold standard was created. For each noun and adjective, 100-300 sentences from the BNC (an average of 267 per word) were randomly extracted. The resulting c. 16K sentences were then manually associated with appropriate SCFs, and the SCF frequency counts were recorded.

To alleviate the manual analysis we developed a tool which first uses the RASP parser with some heuristics to reduce the number of SCF presented, and then allows an annotator to select the preferred choice in a window. The heuristics reduced the average number of SCFs presented alongside each sentence from 52 to 7. The annotator was also presented with an example sentence of each SCF and an intuitive name for the frame, such as PRED (e.g. *Kim is silly*). The program includes an option to record that particular sentences could not (initially) be classified. A screenshot of the tool is shown in Figure 5.

The manual analysis was done by two linguists; one who did the first annotation for the whole data, and another who re-evaluated and corrected some of the initial frame assignments, and classified most of the data left unclassified by the first annotator[2]). A total of 27 SCF types were found for the nouns and 30 for the adjectives in the annotated data. The average number of SCFs taken by nouns was 9 (with the average of 2 added from dictionaries to supplement the manual annotation) and by adjectives 11 (3 of which were from dictionaries). The latter are rare and may not be exemplified in the data given the extraction system.

3.3 Evaluation Measures

We used the standard evaluation metrics to evaluate the accuracy of the SCF lexicons: type precision (the percentage of SCF types that the system proposes

[2]The process precluded measurements of inter-annotator agreement, but this was judged less important than the enhanced accuracy of the gold standard data.

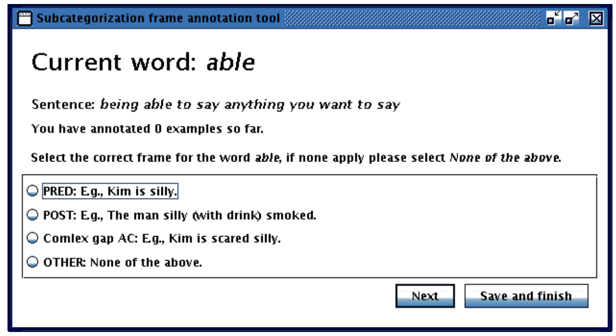

Figure 5: Sample screen of the annotation tool

which are correct), type recall (the percentage of SCF types in the gold standard that the system proposes) and the F-measure which is the harmonic mean of type precision and recall.

We also compared the similarity between the acquired unfiltered[3] SCF distributions and gold standard SCF distributions using various measures of distributional similarity: the Spearman rank correlation (RC), Kullback-Leibler distance (KL), Jensen-Shannon divergence (JS), cross entropy (CE), skew divergence (SD) and intersection (IS). The details of these measures and their application to subcategorization acquisition can be found in (Korhonen and Krymolowski, 2002).

Finally, we recorded the total number of gold standard SCFs unseen in the system output, i.e. the type of false negatives which were never detected by the classifier.

3.4 Results

Table 1 includes the average results for the 183 verbs. The first column shows the results for Briscoe and Carroll's (1997) (B&C) system when this system is run with the original classifier but a more recent version of the parser (Briscoe and Carroll, 2002) and the same filtering technique as our new system (thresholding based on the relative frequencies of SCFs). The classifier of B&C system is comparable to our classifier in the sense that it targets almost the same set of verbal SCFs (165 out of the 168; the 3 additional ones are infrequent in language and thus unlikely to affect the comparison). The second column shows the results for our new system (New).

[3]No threshold was applied to remove the noisy SCFs from the distributions.

Measures	Verbs - Method	
	B&C	New
Precision (%)	47.3	81.8
Recall (%)	40.4	59.5
F-measure	43.6	68.9
KL	3.24	1.57
JS	0.20	0.11
CE	4.85	3.10
SD	1.39	0.74
RC	0.33	0.66
IS	0.49	0.76
Unseen SCFs	28	17

Table 1: Average results for verbs

Measures	Nouns	Adjectives
Precision (%)	91.2	95.5
Recall (%)	47.2	57.6
F-measure	62.2	71.9
KL	0.91	0.69
JS	0.09	0.05
CE	2.03	2.01
SD	0.48	0.36
RC	0.70	0.77
IS	0.62	0.72
Unseen SCFs	15	7

Table 2: Average results for nouns and adjectives

The figures show that the new system clearly performs better than the B&C system. It yields 68.9 F-measure which is a 25.3 absolute improvement over the B&C system. The better performance can be observed on all measures, but particularly on SCF type precision (81.8% with our system vs. 47.3% with the B&C system) and on measures of distributional similarity. The clearly higher IS (0.76 vs. 0.49) and the fewer gold standard SCFs unseen in the output of the classifier (17 vs. 28) indicate that the new system is capable of detecting a higher number of SCFs.

The main reason for better performance is the ability of the new system to detect a number of challenging or complex SCFs which the B&C system could not detect[4]. The improvement is partly attributable to more accurate parses produced by the second release of RASP and partly to the improved SCF classifier developed here. For example, the new system is now able to distinguish predicative PP arguments, such as *I sent him as a messenger* from the wider class of referential PP arguments, supporting discrimination of several syntactically similar SCFs with distinct semantics.

Running our system on the adjective and noun test data yielded the results summarized in Table 2. The F-measure is lower for nouns (62.2) than for verbs (68.9); for adjectives it is slightly better (71.9).[5]

The noun and adjective classifiers yield very high precision compared to recall. The lower recall figures are mostly due to the higher number of gold standard SCFs unseen in the classifier output (rather than, for example, the filtering step). This is particularly evident for nouns for which 15 of the 27 frames exemplified in the gold standard are missing in the classifier output. For adjectives only 7 of the 30 gold standard SCFs are unseen, resulting in better recall (57.6% vs. 47.2% for nouns).

For verbs, subcategorization acquisition performance often correlates with the size of the input data to acquisition (the more data, the better performance). When considering the F-measure results for the individual words shown in Table 3 there appears to be little such correlation for nouns and adjectives. For example, although there are individual high frequency nouns with high performance (e.g. *plan*, freq. 5046, F 90.9) and low frequency nouns with low performance (e.g. *characterisation*, freq. 91, F 40.0), there are also many nouns which contradict the trend (compare e.g. *answer*, freq. 2510, F 50.0 with *fondness*, freq. 71, F 85.7).[6]

Although the SCF distributions for nouns and adjectives appear Zipfian (i.e. the most frequent frames are highly probable, but most frames are infrequent), the total number of SCFs per word is typically smaller than for verbs, resulting in better resistance to sparse data problems.

There is, however, a clear correlation between the performance and the type of gold standard SCFs taken by individual words. Many of the gold stan-

[4]The results reported here for the B&C system are lower than those recently reported in (Korhonen et al., 2006) for the same set of 183 test verbs. This is because we use an improved gold standard. However, the results for the B&C system reported using the less ambitious gold standard are still less accurate (58.6 F-measure) than the ones reported here for the new system.

[5]The results for different word classes are not directly comparable because they are affected by the total number of SCFs evaluated for each word class, which is higher for verbs and lower for nouns and adjectives. This particularly applies to the sensitive measures of distributional similarity.

[6]The frequencies here refer to the number of citations successfully processed by the parser and the classifier.

Noun	F	Adjective	F
abundance	75.0	*able*	66.7
acknowledgement	47.1	*angry*	62.5
answer	50.0	*anxious*	82.4
anxiety	53.3	*aware*	87.5
apology	50.0	*certain*	73.7
appearance	46.2	*clear*	77.8
appointment	66.7	*curious*	57.1
belief	76.9	*desperate*	83.3
call	58.8	*difficult*	77.8
characterisation	40.0	*doubtful*	63.6
communication	40.0	*eager*	83.3
condition	66.7	*easy*	66.7
danger	76.9	*generous*	57.1
decision	70.6	*imperative*	81.8
definition	42.8	*important*	60.9
demand	66.7	*impractical*	71.4
desire	71.4	*improbable*	54.6
doubt	66.7	*insistent*	80.0
evidence	66.7	*kind*	66.7
examination	54.6	*likely*	66.7
experimentation	60.0	*practical*	88.9
fondness	85.7	*probable*	80.0
message	66.7	*sure*	84.2
obsession	54.6	*unaware*	85.7
plan	90.9	*uncertain*	60.0
provision	70.6	*unclear*	63.2
reminder	63.2	*unimportant*	61.5
rumour	61.5	*unlikely*	69.6
temptation	71.4	*unspecified*	50.0
use	60.0	*unsure*	90.0

Table 3: System performance for each test noun and adjective

dard nominal and adjectival SCFs unseen by the classifier involve complex complementation patterns which are challenging to extract, e.g. those exemplified in *The* **argument** *of Jo with Kim about Fido surfaced*, *Jo's* **preference** *that Kim be sacked surfaced*, and *that Sandy came is* **certain**. In addition, many of these SCFs unseen in the data are also very low in frequency, and some may even be true negatives (recall that the gold standard was supplemented with additional SCFs from dictionaries, which may not necessarily appear in the test data).

The main problem is that the RASP parser systematically fails to select the correct analysis for some SCFs with nouns and adjectives regardless of their context of occurrence. In future work, we hope to alleviate this problem by using the weighted GR output from the top *n*-ranked parses returned by the parser as input to the SCF classifier.

4 Discussion

The current system needs refinement to alleviate the bias against some SCFs introduced by the parser's unlexicalized parse selection model. We plan to investigate using weighted GR output with the classifier rather than just the GR set from the highest ranked parse. Some SCF classes also need to be further resolved mainly to differentiate control options with predicative complementation. This requires a lexico-semantic classification of predicate classes.

Experiments with Briscoe and Carroll's system have shown that it is possible to incorporate some semantic information in the acquisition process using a technique that smooths the acquired SCF distributions using back-off (i.e. probability) estimates based on lexical-semantic classes of verbs (Korhonen, 2002). The estimates help to correct the acquired SCF distributions and predict SCFs which are rare or unseen e.g. due to sparse data. They could also form the basis for predicting control of predicative complements.

We plan to modify and extend this technique for the new system and use it to improve the performance further. The technique has so far been applied to verbs only, but it can also be applied to nouns and adjectives because they can also be classified on lexical-semantic grounds. For example, the adjective *simple* belongs to the class of EASY adjectives, and this knowledge can help to predict that it takes similar SCFs to the other class members and that control of 'understood' arguments will pattern with *easy* (e.g. *easy, difficult, convenient*): *The problem will be simple for John to solve*, *For John to solve the problem will be simple*, *The problem will be simple to solve*, etc.

Further research is needed before highly accurate lexicons encoding information also about semantic aspects of subcategorization (e.g. different predicate senses, the mapping from syntactic arguments to semantic representation of argument structure, selectional preferences on argument heads, diathesis alternations, etc.) can be obtained automatically. However, with the extensions suggested above, the system presented here is sufficiently accurate for building an extensive SCF lexicon capable of supporting various NLP application tasks. Such a lexicon will be built and distributed for research purposes along with the gold standard described here.

5 Conclusion

We have described the first system for automatically acquiring verbal, nominal and adjectival subcategorization and associated frequency information from English corpora, which can be used to build large-scale lexicons for NLP purposes. We have also described a new annotation tool for producing training and test data for the task. The acquisition system, which is capable of distinguishing 168 verbal, 37 adjectival and 31 nominal frames, classifies corpus occurrences to SCFs on the basis of GRs produced by a robust statistical parser. The information provided by GRs closely matches the structure that subcategorization acquisition seeks to recover. Our experiment shows that the system achieves state-of-the-art performance with each word class. The discussion suggests ways in which we could improve the system further before using it to build a large subcategorization lexicon capable of supporting various NLP application tasks.

Acknowledgements

This work was supported by the Royal Society and UK EPSRC project 'Accurate and Comprehensive Lexical Classification for Natural Language Processing Applications' (ACLEX). We would like to thank Diane Nicholls for her help during this work.

References

B. Boguraev, J. Carroll, E. J. Briscoe, D. Carter, and C. Grover. 1987. The derivation of a grammatically-indexed lexicon from the Longman Dictionary of Contemporary English. In *Proc. of the 25th Annual Meeting of ACL*, pages 193–200, Stanford, CA.

M. Brent. 1991. Automatic acquisition of subcategorization frames from untagged text. In *Proc. of the 29th Meeting of ACL*, pages 209–214.

E. J. Briscoe and J. Carroll. 1997. Automatic Extraction of Subcategorization from Corpora. In *Proc. of the 5th ANLP*, Washington DC, USA.

E. J. Briscoe and J. Carroll. 2002. Robust accurate statistical annotation of general text. In *Proc. of the 3rd LREC*, pages 1499–1504, Las Palmas, Canary Islands, May.

E. J. Briscoe, J. Carroll, and R. Watson. 2006. The second release of the rasp system. In *Proc. of the COLING/ACL 2006 Interactive Presentation Sessions*, Sydney, Australia.

L. Burnard, 1995. *The BNC Users Reference Guide*. British National Corpus Consortium, Oxford, May.

G. Carroll and M. Rooth. 1998. Valence induction with a head-lexicalized pcfg. In *Proc. of the 3rd Conference on EMNLP*, Granada, Spain.

J. Carroll, G. Minnen, and E. J. Briscoe. 1998. Can Subcategorisation Probabilities Help a Statistical Parser? In *Proceedings of the 6th ACL/SIGDAT Workshop on Very Large Corpora*, pages 118–126, Montreal, Canada.

R. Grishman, C. Macleod, and A. Meyers. 1994. COMLEX Syntax: Building a Computational Lexicon. In *COLING*, Kyoto.

A. Korhonen and Y. Krymolowski. 2002. On the Robustness of Entropy-Based Similarity Measures in Evaluation of Subcategorization Acquisition Systems. In *Proc. of the Sixth CoNLL*, pages 91–97, Taipei, Taiwan.

A. Korhonen, Y. Krymolowski, and Z. Marx. 2003. Clustering Polysemic Subcategorization Frame Distributions Semantically. In *Proc. of the 41st Annual Meeting of ACL*, pages 64–71, Sapporo, Japan.

A. Korhonen, Y. Krymolowski, and E. J. Briscoe. 2006. A large subcategorization lexicon for natural language processing applications. In *Proc. of the 5th LREC*, Genova, Italy.

A. Korhonen. 2002. *Subcategorization acquisition*. Ph.D. thesis, University of Cambridge Computer Laboratory.

C. Macleod, A. Meyers, R. Grishman, L. Barrett, and R. Reeves. 1997. Designing a dictionary of derived nominals. In *Proc. of RANLP*, Tzigov Chark, Bulgaria.

C. Manning. 1993. Automatic Acquisition of a Large Subcategorization Dictionary from Corpora. In *Proc. of the 31st Meeting of ACL*, pages 235–242.

S. Schulte im Walde and C. Brew. 2002. Inducing german semantic verb classes from purely syntactic subcategorisation information. In *Proc. of the 40th Annual Meeting of ACL*, Philadephia, USA.

M. Surdeanu, S. Harabagiu, J. Williams, and P. Aarseth. 2003. Using predicate-argument structures for information extraction. In *Proc. of the 41st Annual Meeting of ACL*, Sapporo.

J. Yallop, A. Korhonen, and E. J. Briscoe. 2005. Automatic acquisition of adjectival subcategorization from corpora. In *Proc. of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 614–621, Ann Arbor, Michigan.