

Automatic Acquisition of Adjectival Subcategorization from Corpora

Jeremy Yallop*, Anna Korhonen, and Ted Briscoe

Computer Laboratory
University of Cambridge
15 JJ Thomson Avenue
Cambridge CB3 0FD, UK

yallop@cantab.net, {Anna.Korhonen, Ted.Briscoe}@cl.cam.ac.uk

Abstract

This paper describes a novel system for acquiring adjectival subcategorization frames (SCFs) and associated frequency information from English corpus data. The system incorporates a decision-tree classifier for 30 SCF types which tests for the presence of grammatical relations (GRs) in the output of a robust statistical parser. It uses a powerful pattern-matching language to classify GRs into frames hierarchically in a way that mirrors inheritance-based lexica. The experiments show that the system is able to detect SCF types with 70% precision and 66% recall rate. A new tool for linguistic annotation of SCFs in corpus data is also introduced which can considerably alleviate the process of obtaining training and test data for subcategorization acquisition.

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

Part of this research was conducted while this author was at the University of Edinburgh Laboratory for Foundations of Computer Science.

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., 1998b), 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 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., 1998b; Korhonen et al., 2003)), besides systems for a number of other languages (e.g. (Kawahara and Kurohashi, 2002; Ferrer, 2004)).

While there has been considerable research into acquisition of verb subcategorization, we are not aware of any systems built for adjectives. Although adjectives are syntactically less multivalent than verbs, and although verb subcategorization distribution data appears to offer the greatest potential boost in parser performance, accurate and comprehensive knowledge of the many adjective SCFs can improve the accuracy of parsing at several levels

(from tagging to syntactic and semantic analysis). Automatic SCF acquisition techniques are particularly important for adjectives because extant syntax dictionaries provide very limited coverage of adjective subcategorization.

In this paper we propose a method for automatic acquisition of adjectival SCFs from English corpus data. Our method has been implemented using a decision-tree classifier which tests for the presence of grammatical relations (GRs) in the output of the RASP (Robust Accurate Statistical Parsing) system (Briscoe and Carroll, 2002). It uses a powerful task-specific pattern-matching language which enables the frames to be classified hierarchically in a way that mirrors inheritance-based lexica. As reported later, the system is capable of detecting 30 SCFs with an accuracy comparable to that of best state-of-art verbal SCF acquisition systems (e.g. (Korhonen, 2002)).

Additionally, we present a novel 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 discuss adjectival subcategorization in section 2 and introduce the system for SCF acquisition in section 3. Details of the annotation tool and the experimental evaluation are supplied in section 4. Section 5 provides discussion on our results and future work, and section 6 summarises the paper.

2 Adjectival Subcategorization

Although the number of SCF types for adjectives is smaller than the number reported for verbs (e.g. (Briscoe and Carroll, 1997)), adjectives nevertheless exhibit rich syntactic behaviour. Besides the common attributive and predicative positions there are at least six further positions in which adjectives commonly occur (see figure 1). Adjectives in predicative position can be further classified according to the nature of the arguments with which they combine — finite and non-finite clauses and noun phrases, phrases with and without complementisers, etc. — and whether they occur as subject or object. Additional distinctions can be made concern-

Attributive	“The young man”
Predicative	“He is young”
Postpositive	“Anyone [who is] young can do it”
Predeterminer	“such a young man”; “so young a man”
Fused modifier-head	“the younger of them”; “the young”
Predicative adjunct	“he died young”
Supplementive clause	“Young, he was plain in appearance”
Contingent clause	“When young, he was lonely”

Figure 1: Fundamental adjectival frames

ing such features as the mood of the complement (mandative, interrogative, etc.), preferences for particular prepositions and whether the subject is extraposed.

Even ignoring preposition preference, there are more than 30 distinguishable adjectival SCFs. Some fairly extensive frame sets can be found in large syntax dictionaries, such as COMLEX (31 SCFs) (Wolff et al., 1998) and ANLT (24 SCFs) (Boguraev et al., 1987). While such resources are generally accurate, they are disappointingly incomplete: none of the proposed frame sets in the well-known resources subsumes the others, the coverage of SCF types for individual adjectives is low, and (accurate) information on the relative frequency of SCFs for each adjective is absent.

The inadequacy of manually-created dictionaries and the difficulty of adequately enhancing and maintaining the information by hand was a central motivation for early research into automatic subcategorization acquisition. The focus heretofore has remained firmly on verb subcategorization, but this is not sufficient, as countless examples show. Knowledge of adjectival subcategorization can yield further improvements in tagging (e.g. distinguishing between “to” as an infinitive marker and as a true preposition), parsing (e.g. distinguishing between PP-arguments and adjuncts), and semantic analysis. For example, if John is both *easy* and *eager* to please then *we* know that he is the recipient of pleasure in the first instance and desirous of providing it in the second, but a computational system cannot determine this without knowledge of the subcategorization of the two adjectives. Likewise, a natural language generation system can legitimately apply the extraposition transformation to the first case, but not to the second: It is “easy to please John”, but not

“eager” to do so, at least if “it” be expletive. Similar examples abound.

Many of the difficulties described in the literature on acquiring verb subcategorization also arise in the adjectival case. The most apparent is data sparsity: among the 100M-word British National Corpus (BNC) (Burnard, 1995), the RASP tools find 124,120 distinct adjectives, of which 70,246 occur only once, 106,464 fewer than ten times and 119,337 fewer than a hundred times. There are fewer than 1,000 adjectives in the corpus which have more than 1,000 occurrences. Both adjective and SCF frequencies have Zipfian distributions; consequently, even the largest corpora may contain only single instances of a particular adjective-SCF combination, which is generally insufficient for classification.

3 Description of the System

Besides focusing on adjectives, our approach to SCF acquisition differs from earlier work in a number of ways. A common strategy in existing 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 the patterns are extracted from GRs — representations of head-complement relations which are designed to be largely parser-independent — making the techniques more widely applicable and allowing classification to operate at a higher level. Further, most existing systems work by classifying corpus occurrences into individual, mutually independent SCFs. We adopt instead a hierarchical approach, viewing frames that share features as descendants of a common parent frame. The benefits are severalfold: specifying each feature only once makes the system both more efficient and easier to understand and maintain, and the multiple inheritance hierarchy reflects the hierarchy of lexical types found in modern grammars where relationships between similar frames are represented explicitly¹.

Our acquisition process consists of two main steps: 1) extracting GRs from corpus data, and 2) feeding the GRs as input to the classifier which incrementally matches parts of the GR sets to decide which branches of a decision-tree to follow. The

¹Compare the cogent argument for an inheritance-based lexicon in (Flickinger and Nerbonne, 1992), much of which can be applied unchanged to the taxonomy of SCFs.

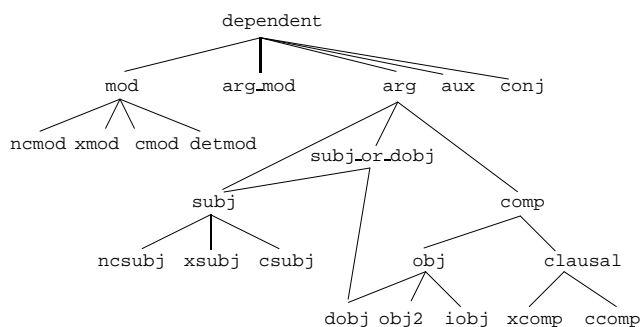


Figure 2: The GR hierarchy used by RASP

leaves of the tree correspond to SCFs. The details of these two steps are provided in the subsequent sections, respectively².

3.1 Obtaining Grammatical Relations

Attempts to acquire verb subcategorization have benefited from increasingly sophisticated parsers. We have made use of the RASP toolkit (Briscoe and Carroll, 2002) — a modular statistical parsing system which includes a tokenizer, tagger, lemmatiser, and a wide-coverage unification-based tag-sequence parser. The parser has several modes of operation; we invoked it in a mode in which GRs with associated probabilities are emitted even when a complete analysis of the sentence could not be found. In this mode there is wide coverage (over 98% of the BNC receives at least a partial analysis (Carroll and Briscoe, 2002)) which is useful in view of the infrequent occurrence of some of the SCFs, although combining the results of competing parses may in some cases result in an inconsistent or misleading combination of GRs.

The parser uses a scheme of GRs between lemmatised lexical heads (Carroll et al., 1998a; Briscoe et al., 2002). The relations are organized as a multiple-inheritance subsumption hierarchy where each subrelation extends the meaning, and perhaps the argument structure, of its parents (figure 2). For descriptions and examples of each relation, see (Carroll et al., 1998a).

The dependency relationships which the GRs embody correspond closely to the head-complement

²In contrast to almost all earlier work, there was no filtering stage involved in SCF acquisition. The classifier was designed to operate with high precision, so filtering was less necessary.

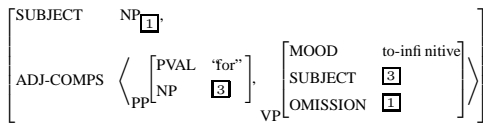


Figure 3: Feature structure for SCF adj-obj-for-to-inf

```
( |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])
xmod(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 4: GRs from RASP for adj-obj-for-to-inf

structure which subcategorization acquisition attempts to recover, which makes GRs ideal input to the SCF classifier. Consider the arguments of “easy” 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 3, (using AVM notation in place of COMLEX s-expressions). Part of the output of RASP for this sentence (the full output includes 87 weighted GRs) is shown in figure 4³.

Each instantiated GR in figure 4 corresponds to one or more parts of the feature structure in figure 3. `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 `xmod(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 [3]: the subject of the

³The format is slightly more complicated than that shown in (Carroll et al., 1998a): each argument that corresponds to a word consists of three parts: the lexeme, the part of speech tag, and the position (index) of the word in the sentence.

```
xcomp(_ , [*;1;be-verb], ~)
xmod([to;*;to], 1, [*;2;vv0])
ncsubj(1, [*;3;noun/pronoun], _)
ncmod([for;*;if], ~, [*;4;noun/pronoun])
ncsubj(2, 4)
```

Figure 5: A pattern to match the frame adj-obj-for-to-inf

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 [1] of the VP-complement and the subject of the whole clause.

3.2 SCF Classifier

3.2.1 SCF Frames

We used for our classifier a modified version of the fairly extensive COMLEX frameset, including 30 SCFs. The COMLEX frameset includes mutually inconsistent frames, such as sentential complement with obligatory complementiser *that* and sentential complement with optional *that*. We modified the frameset so that an adjective can legitimately instantiate any combination of frames, which simplifies classification. We also added simple-predicative and attributive SCFs to the set, since these account for a substantial proportion of frame instances. Finally, frames which could only be distinguished by information not retained in the GRs scheme of the current version of the shallow parser were merged (e.g. the COMLEX frames `adj-subj-to-inf-rs` (“She was kind to invite me”) and `adj-to-inf` (“She was able to climb the mountain”).

3.2.2 Classifier

The classifier operates by attempting to match the set of GRs associated with each sentence against various patterns. The patterns were developed by a combination of knowledge of the GRs and examining a set of training sentences to determine which relations were actually emitted by the parser for each SCF. The data used during development consisted of the sentences in the BNC in which one of the 23 adjectives⁴ given as examples for SCFs in (Macleod

⁴The adjectives used for training were: *able, anxious, apparent, certain, convenient, curious, desirable, disappointed, easy, happy, helpful, imperative, impractical, insistent, kind, obvious, practical, preferable, probable, ridiculous, unaware, uncertain* and *unclear*.

et al., 1998) occur.

In our pattern matching language a pattern is a disjunction of sets of partially instantiated GRs with logic variables (*slots*) in place of indices, augmented by ordering constraints that restrict the possible instantiations of slots. A match is considered successful if the set of GRs can be unified with any of the disjuncts. Unification of a sentence-relation and a pattern-relation occurs when there is a one-to-one correspondence between sentence elements and pattern elements that includes a mapping from slots to indices (a *substitution*), and where atomic elements in corresponding positions share a common subtype.

Figure 5 shows a pattern for matching the SCF `adj-obj-for-to-inf`. For a match to succeed there must be GRs associated with the sentence that match each part of the pattern. Each argument matches either anything at all (*), the “current” adjective (~), an empty GR argument (.), a `[word;id;part-of-speech]` 3-tuple or a numeric id. In a successful match, equal ids in different parts of the pattern must match the same word position, and distinct ids must match different positions.

The various patterns are arranged in a tree, where a parent node contains the elements common to all of its children. This kind of once-only representation of particular features, together with the successive refinements provided by child nodes reflects the organization of inheritance-based lexica. The inheritance structure naturally involves multiple inheritance, since each frame typically includes multiple features (such as the presence of a *to-infinitive* complement or an expletive subject argument) inherited from abstract parent classes, and each feature is instantiated in several frames.

The tree structure also improves the efficiency of the pattern matching process, which then occurs in stages: at each matching node the classifier attempts to match a set of relations with each child pattern to yield a substitution that subsumes the substitution resulting from the parent match.

Both the patterns and the pattern language itself underwent successive refinements after investigation of the performance on training data made it increasingly clear what sort of distinctions were useful to express. The initial pattern language had no slots; it was easy to understand and implement, but insufficiently expressive. The final refinement was the ad-

unspecified	285	improbable	350
unsure	570	doubtful	1147
generous	2052	sure	13591
difficult	18470	clear	19617
important	33303		

Table 1: Test adjectives and frequencies in the BNC

dition of ordering constraints between instantiated slots, which are indispensable for detecting, e.g., extraposition.

4 Experimental Evaluation

4.1 Data

In order to evaluate the system we selected a set of 9 adjectives which between them could instantiate all of the frames. The test set was intentionally kept fairly small for these first experiments with adjectival SCF acquisition so that we could carry out a thorough evaluation of all the test instances. We excluded the adjectives used during development and adjectives with fewer than 200 instances in the corpus. The final test set, together with their frequencies in the tagged version of the BNC, is shown in table 1. For each adjective we extracted 200 sentences (evenly spaced throughout the BNC) which we processed using the SCF acquisition system described in the previous section.

4.2 Method

4.2.1 Annotation Tool and Gold Standard

Our gold standard was human-annotated data. Two annotators associated a SCF with each sentence/adjective pair in the test data. To alleviate the process we developed a program which first uses reliable heuristics to reduce the number of SCF choices and then allows the annotator to select the preferred choice with a single mouse click in a browser window. The heuristics reduced the average number of SCFs presented alongside each sentence from 30 to 9. Through the same browser interface we provided annotators with information and instructions (with links to COMLEX documentation), the ability to inspect and review previous decisions and decision summaries⁵ and an option to record that partic-

⁵The varying number of SCFs presented to the user and the ability to revisit previous decisions precluded accurate measure-



Figure 6: Sample classification screen for web annotation tool

ular sentences could not be classified (which is useful for further system development, as discussed in section 5). A screenshot is shown in figure 6. The resulting annotation revealed 19 of the 30 SCFs in the test data.

4.2.2 Evaluation Measures

We use the standard evaluation metrics: type and token precision, recall and F-measure. *Token recall* is the proportion of annotated (*sentence, frame*) pairs that the system recovered correctly. *Token precision* is the proportion of classified (*sentence, frame*) pairs that were correct. *Type precision* and *type recall* are analogously defined for (*adjective, frame*) pairs. The F-measure ($\beta = 1$) is a weighted combination of precision and recall.

4.3 Results

Running the system on the test data yielded the results summarised in table 2. The greater expressiveness of the final pattern language resulted in a classifier that performed better than the ‘regression’ versions which ignored either ordering constraints, or both ordering constraints and slots. As expected, removing features from the classifier translated directly into degraded accuracy. The performance of the best classifier (67.8% F-measure) is quite similar to that of the best current verbal SCF acquisition systems (e.g. (Korhonen, 2002)).

Results for individual adjectives are given in table 3. The first column shows the number of SCFs acquired for each adjective, ranging from 2 for *unspec-*

ments of inter-annotator agreement, but this was judged less important than the enhanced ease of use arising from the reduced set of choices.

Type performance

System	Precision	Recall	F
Final	69.6	66.1	67.8
No order constraints	67.3	62.7	64.9
No slots	62.7	51.4	56.5

Token performance

System	Precision	Recall	F
Final	63.0	70.5	66.5
No order constraints	58.8	68.3	63.2
No slots	58.3	67.6	62.6

Table 2: Overall performance of the classifier and of regression systems with restricted pattern-matching

ified to 11 for *doubtful*. Looking at the F-measure, the best performing adjectives are *unspecified*, *difficult* and *sure* (80%) and the worst performing *unsure* (50%) and *improbable* (60%).

There appears to be no obvious connection between performance figures and the number of acquired SCF types; differences are rather due to the difficulty of detecting individual SCF types — an issue directly related to data sparsity.

Despite the size of the BNC, 5 SCFs were not seen at all, either for the test adjectives or for any others. Frames involving *to-infinitive* complements were particularly rare: 4 such SCFs had no examples in the corpus and a further 3 occurred 5 times or fewer in the test data. It is more difficult to develop patterns for SCFs that occur infrequently, and the few instances of such SCFs are unlikely to include a set of GRs that is adequate for classification. The effect on the results was clear: of the 9 SCFs which the classifier did not correctly recognise at all, 4 occurred 5 times or fewer in the test data and a further 2 occurred 5–10 times.

The most common error made by the classifier was to mistake a complex frame (e.g. *adj-obj-for-to-inf*, or *to-inf-wh-adj*) for *simple-predicative*, which subsumes all such frames. This occurred whenever the GRs emitted by the parser failed to include any information about the complements of the adjective.

5 Discussion

Data sparsity is perhaps the greatest hindrance both to recovering adjectival subcategorization and to lexical acquisition in general. In the future, we plan to carry out experiments with a larger set of adjec-

Adjective	SCFs	Precision	Recall	F-measure
unspecified	2	66.7	100.0	80.0
generous	3	60.0	100.0	75.0
improbable	5	60.0	60.0	60.0
unsure	6	50.0	50.0	50.0
important	7	55.6	71.4	62.5
clear	8	83.3	62.5	71.4
difficult	8	85.7	75.0	80.0
sure	9	100.0	66.7	80.0
doubtful	11	66.7	54.5	60.0

Table 3: SCF count and classifier performance for each adjective.

tives using more data (possibly from several corpora and the web) to determine how severe this problem is for adjectives. One possible way to address the problem is to smooth the acquired SCF distributions using SCF “back-off” (probability) estimates based on lexical classes of adjectives in the manner proposed by (Korhonen, 2002). This helps to correct the acquired distributions and to detect low frequency and unseen SCFs.

However, our experiment also revealed other problems which require attention in the future. One such is that GRS output by RASP (the version we used in our experiments) do not retain certain distinctions which are essential for distinguishing particular SCFs. For example, a sentential complement of an adjective with a *that*-complementiser should be annotated with `ccomp(that, adjective, verbal-head)`, but this relation (with `that` as the *type* argument) does not occur in the parsed BNC. As a consequence the classifier is unable to distinguish the frame.

Another problem arises from the fact that our current classifier operates on a predefined set of SCFs. The COMLEX SCFs, from which ours were derived, are extremely incomplete. Almost a quarter (477 of 1931) of sentences were annotated as “undefined”. For example, while there are SCFs for sentential and infinitival complement in subject position with *what*⁶, there is no SCF for the case with a *what*-prefixed complement in object position, where the subject is an NP. The lack is especially perplexing, because COMLEX does include the corresponding SCFs for verbs. There is a frame for “He wondered

⁶(adj-subj-what-s: “What he will do is uncertain”; adj-subj-what-to-inf: “What to do was unclear”), together with the extraposed versions (extrap-adj-what-s and extrap-adj-what-to-inf).

what to do” (`what-to-inf`), but none for “He was unsure what to do”.

While we can easily extend the current frame-set by looking for further SCF types from dictionaries and from among the corpus occurrences labelled by our annotators as unclassified, we also plan to extend the classifier to automatically induce previously unseen frames from data. A possible approach is to use restricted generalization on sets of GRS to group similar sentences together. Generalization (*anti-unification*) is an intersection operation on two structures which retains the features common to both; generalization over the sets of GRS associated with the sentences which instantiate a particular frame can produce a pattern such as we used for classification in the experiments described above. This approach also offers the possibility of associating confidence levels with each pattern, corresponding to the degree to which the generalized pattern captures the features common to the members of the associated class. It is possible that frames could be induced by grouping sentences according to the “best” (e.g. most information-preserving) generalizations for various combinations, but it is not clear how this can be implemented with acceptable efficiency.

The hierarchical approach described in this paper may also be helpful in the discovery of new frames: missing combinations of parent classes can be explored readily, and it may be possible to combine the various features in an SCF feature structure to generate example sentences which a human could then inspect to judge grammaticality.

6 Conclusion

We have described a novel system for automatically acquiring adjectival subcategorization and associated frequency information from corpora, along with an annotation tool for producing training and test data for the task. The acquisition system, which is capable of distinguishing 30 SCF types, performs sophisticated pattern matching on sets of GRS produced by a robust statistical parser. The information provided by GRS closely matches the structure that subcategorization acquisition seeks to recover. The figures reported demonstrate the feasibility of the approach: our classifier achieved 70% type pre-

cision and 66% type recall on the test data. The discussion suggests several ways in which the system may be improved, refined and extended in the future.

Acknowledgements

We would like to thank Ann Copestake for all her help during this work.

References

- B. Boguraev, J. Carroll, E. Briscoe, D. Carter, and C. Grover. 1987. The derivation of a grammatically-indexed lexicon from the Longman Dictionary of Contemporary English. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, pages 193–200, Stanford, CA.
- Michael R. Brent. 1991. Automatic acquisition of subcategorization frames from untagged text. In *Meeting of the Association for Computational Linguistics*, pages 209–214.
- E. J. Briscoe and J. Carroll. 1997. Automatic Extraction of Subcategorization from Corpora. In *Proceedings of the 5th Conference on Applied Natural Language Processing*, Washington DC, USA.
- E. Briscoe and J. Carroll. 2002. Robust accurate statistical annotation of general text. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, pages 1499–1504, Las Palmas, Canary Islands, May.
- E. Briscoe, J. Carroll, Jonathan Graham, and Ann Copestake. 2002. Relational evaluation schemes. In *Proceedings of the Beyond PARSEVAL Workshop at the 3rd International Conference on Language Resources and Evaluation*, pages 4–8, Las Palmas, Gran Canaria.
- Lou Burnard, 1995. *The BNC Users Reference Guide*. British National Corpus Consortium, Oxford, May.
- J. Carroll and E. Briscoe. 2002. High precision extraction of grammatical relations. In *Proceedings of the 19th International Conference on Computational Linguistics*, pages 134–140, Taipei, Taiwan.
- Glenn Carroll and Mats Rooth. 1998. Valence induction with a head-lexicalized pcfg. In *Proc. of the 3rd Conference on Empirical Methods in Natural Language Processing*, Granada, Spain.
- J. Carroll, E. Briscoe, and A. Sanfilippo. 1998a. Parser evaluation: a survey and a new proposal. In *Proceedings of the 1st International Conference on Language Resources and Evaluation*, pages 447–454, Granada, Spain.
- John Carroll, Guido Minnen, and Edward Briscoe. 1998b. Can Subcategorisation Probabilities Help a Statistical Parser? In *Proceedings of the 6th ACL/SIGDAT Workshop on Very Large Corpora*, pages 118–126, Montreal, Canada. Association for Computational Linguistics.
- Eva Esteve Ferrer. 2004. Towards a Semantic Classification of Spanish Verbs Based on Subcategorisation Information. In *ACL Student Research Workshop*, Barcelona, Spain.
- Dan Flickinger and John Nerbonne. 1992. Inheritance and complementation: A case study of easy adjectives and related nouns. *Computational Linguistics*, 18(3):269–309.
- Daisuke Kawahara and Sadao Kurohashi. 2002. Fertilization of Case Frame Dictionary for Robust Japanese Case Analysis. In *19th International Conference on Computational Linguistics*.
- Anna Korhonen, Yuval Krymolowski, and Zvika Marx. 2003. Clustering Polysemic Subcategorization Frame Distributions Semantically. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 64–71, Sapporo, Japan.
- Anna Korhonen. 2002. *Subcategorization acquisition*. Ph.D. thesis, University of Cambridge Computer Laboratory, February.
- Catherine Macleod, Ralph Grishman, and Adam Meyers, 1998. *COMLEX Syntax Reference Manual*. Computer Science Department, New York University.
- Christopher D. Manning. 1993. Automatic Acquisition of a Large Subcategorization Dictionary from Corpora. In *Meeting of the Association for Computational Linguistics*, pages 235–242.
- S. Schulte im Walde and C. Brew. 2002. Inducing german semantic verb classes from purely syntactic subcategorisation information. In *40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, USA.
- Mihai Surdeanu, Sanda Harabagiu, John Williams, and Paul Aarseth. 2003. Using predicate-argument structures for information extraction. In *Proc. of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo.
- Susanne Rohen Wolff, Catherine Macleod, and Adam Meyers, 1998. *COMLEX Word Classes Manual*. Computer Science Department, New York University, June.