

Learning Argument/Adjunct Distinction for Basque

**Izaskun Aldezabal, Maxux Aranzabe, Koldo
Gojenola, Kepa Sarasola**

Dept. of Computer Languages and Systems,
University of the Basque Country, 649 P. K.,
E-20080 Donostia,
Basque Country

Aitziber Atutxa

University of Maryland
College Park
Maryland, 20740
jibatsaa@si.ehu.es

Abstract

This paper presents experiments performed on lexical knowledge acquisition in the form of verbal argumental information. The system obtains the data from raw corpora after the application of a partial parser and statistical filters. We used two different statistical filters to acquire the argumental information: Mutual Information, and Fisher's Exact test.

Due to the characteristics of agglutinative languages like Basque, the usual classification of arguments in terms of their syntactic category (such as NP or PP) is not suitable. For that reason, the arguments will be classified in 48 different kinds of case markers, which makes the system fine grained if compared to equivalent systems that have been developed for other languages.

This work addresses the problem of distinguishing arguments from adjuncts, this being one of the most significant sources of noise in subcategorization frame acquisition.

Introduction

In recent years a considerable effort has been done on the acquisition of lexical information. As several authors point out, this information is useful for a wide range of applications. For example, J. Carroll et al. (1998) show how adding subcategorization information improves the performance of a parser.

With this in mind our aim is to obtain a system that automatically discriminates between subcategorized elements of verbs (arguments) and non-subcategorized ones (adjuncts).

We have evaluated our system in two ways: comparing the results to a gold standard and estimating the coverage over sentences in the

corpus. The purpose was to find out which was the impact of each approach on this particular task. The two methods of evaluation yield significantly different results.

Basque is the subject of this study. A language that, in contrast to languages like English, has limited resources in the form of digital corpora, computational lexicons, grammars or annotated treebanks. Therefore, any effort like the one presented here, oriented to create lexical resources, has to be driven to do as much automatic work as possible, minimizing development costs.

The paper is divided into 4 sections. The first section is devoted to explain the theoretical motivations underlying the process. The second section is a description of the different stages of the system. The third section presents the results obtained. The fourth section is a review of previous work on automatic subcategorization acquisition. Finally, we present the main conclusions.

1 The argument/adjunct distinction

Talking about Subcategorization Frames (SCF), means talking about arguments. Many existing systems acquire directly a set of possible SCFs without any previous filtering of adjuncts. However, adjuncts are a substantial source of noise and therefore, in order to avoid this problem, our approach addresses the problem of the argument/adjunct distinction.

The argument/adjunct distinction is probably one of the most unclear issues in linguistics. The distinction has been presented, for example, in the generativist tradition, in the following way: arguments are those elements participating in the event and adjuncts are those elements contextualizing or locating the event.

This definition seems to be quite clear, but when we deal with concrete examples it is not the

case. For example, if we take two verbs, *talk* and *play*.

- a. Yesterday I talked with Mary.
- b. Yesterday I played soccer with Mary.

Here *Mary* is a participant of the event in both cases, therefore under the given definition both would be arguments. But this is contradictory to what traditional views consider in practice. The PP, *with Mary*, is considered an argument of *talk* but not an argument of *play*. It is true that there are differences between both of them because *playing* does not require two participants (though it can have them), while *talking* (under the sense of communicating) seems to require two participants.

Finer argument/adjunct distinction have also been proposed differentiating between basic arguments, pseudo-arguments and adjuncts. Basic arguments are those required by the verb. Pseudo-arguments are those that even if they are not required by the verb, when appearing they extend the verbal semantics, for example, adding new participants. And finally adjuncts, which would be contextualizers of the event. The most radical view is to consider the argument/adjunct distinction as a continuum where the elements belonging to the extremes of this continuum can be easily classified as arguments or adjuncts. On the contrary, the elements belonging to the central part of the continuum can be easily misclassified. For further reference see C. Schutze (1995), J.M. Gawron (1986), C. Verspoor (1997), J. Grimshaw (1990), and N. Chomsky (1995).

From the different diagnostics proposed in the literature some are quite consistent among various authors (R. Grishman et al. 1994, C. Pollard and I. Sag 1987, C. Verspoor 1997).

- 1) The Obligatoriness condition. When a verb demands obligatorily the appearance of an element, this element will be an argument.
 - a. John put the book *on the table*
 - b. *John put the book
- 2) Frequency. Arguments of a verb occur more frequently with that verb than with the other verbs.
 - a. I came *from home* (argument).
 - b. I heard it *from you* (adjunct).
- 3) Iterability: Several instances of the same adjunct can appear together with a verb, while

several instances of an argument cannot appear with a verb.

- a. I saw you in Washington, in the Kennedy Center.
 - b. *I saw Alice John (being John and Alice two persons)
- 4) Relative order: Arguments tend to appear closer to the verb than adjuncts.
 - a. I put the book on the table *at three*
 - b. *I put *at three* the book on the table
 - 5) Implicational test: Arguments are semantically implied, even when they are optional.
 - a. I came to your house (from x)
 - b. I heard that (from x)

The third and fourth tests were not very useful to us. Iterability test is quite weak because it seems to rely more on some other semantic notions such as part/whole relation than in the argument/adjunct distinction. For example, sentence 3.a would be grammatical due to semantic plausibility. *The Kennedy Center* is a part of *Washington*, therefore to see somebody in *the Kennedy Center* and see him in *Washington* are not semantically incompatible, so it is plausible to say it. In the case of 3.b *John* is not a part of *Alice* and therefore it is not plausible to see *Alice John*. But for example it is plausible to say *I saw you the hand*. The relative order test is difficult to apply on a language like Basque which is a free word order language.

The first and fifth tests are robust enough to be useful in practice. But only the two first diagnostics can be captured statistically by the application of association measures like Mutual Information. We did not come out with any straightforward way to apply the fifth test computationally.

Before talking about the different measures applied, we will present step by step the whole process we pursued for achieving the argument/adjunct distinction.

2 The acquisition process

Our starting point was a raw newspaper corpus from of 1,337,445 words, where there were instances of 1,412 verbs. From them, we selected 640 verbs as statistically relevant because they appear in more than 10 sentences.

1)... (a) [EEBBetako lehendakariak] (b) [UEko 15 herrialdeetako merkataritza ministroekin]
 (c) [bazkaldu behar zuen] (d) [negoiazioen bilgunean] ...

2) ... the president of the USA had to eat with the ministers of Commerce of 15 countries of the UE in the negotiation center ...

<p>(a) [EEBB-etako lehendakari-a-k] [USA-of president-the-erg.] NP-ergative(president, singular) The president of the USA</p>	<p>(b) [UE-ko 15 herrialde-etako merkataritza ministro-ekin] [UE-of 15 countries-of Commerce ministers-with] PP(with)-commitative(minister, plural) with the ministers of Commerce of 15 countries of the UE</p>
<p>(c) [bazkaldu behar zuen] [to eat had] verb(eat) had to eat</p>	<p>(d) [negoiazio-en bilgune-an] [negotiation-of center-in] PP(in)-inessive(center, singular) in the negotiation center</p>

Figure 1. Example of the output of the shallow parsing phase: 1) Input (in Basque), 2) English translation., Below (c) Verb phrase and (a,b,d) verbal dependents (phrases), and also under the case+head

As we said earlier, our goal was to distinguish arguments from adjuncts. When starting from raw corpus, like in this case, it is necessary to get instances of verbs together with their dependents (arguments and adjuncts). We obtained this information applying a partial parser (section 2.1) to the corpus. Once we had the dependents, statistical measures helped us deciding which were arguments and which were adjuncts (section 2.2).

2.1 The parsing phase

Aiming to obtain the data against which statistical filters will be applied, we analyzed the corpus using several available linguistic resources:

- First, we performed morphological analysis of the corpus, based on two-level morphology (K. Koskeniemi 1983; I. Alegria et al. 1996) and disambiguation using the Constraint Grammar formalism (Karlsson et al. 1995, Aduriz et al. 1997).
- Second, a shallow parser was applied (I. Aldezabal et al. 2000), which recognizes basic syntactic units including noun phrases, prepositional phrases and several types of subordinate sentences.
- The third step consisted in linking each verb and its dependents. Basque lacks a robust parser as in (T. Briscoe & J. Carroll 1997, D. Kawahara et al. 2001) and, therefore, we used a finite state grammar to link the dependents (both arguments and adjuncts) with the verb (I.

I. Aldezabal et al. 2001). This grammar was developed using the Xerox Finite State Tool (L. Karttunen et al. 1997). Figure 1 shows the result of the parsing phase. In this case, both commitative and inessive cases (PPs) are adjuncts, while the ergative NP is an argument. The linking of dependents to a verb is not trivial considering that Basque is a language with free order of constituents, and any element appearing between two verbs could be, in principle, dependent on any of them. Many problems must be taken into account, such as ambiguity and determination of clause boundaries, among others. We evaluated the accuracy up to this point, obtaining a precision over dependents of 87% and a recall of 66%. So the input data to the next phase was relatively noisy.

2.2 The argument selection phase

In the data resulting from the shallow parsing phase we counted up to 65 different cases (types of arguments, including postpositions and different types of suffixes). These are divided in two main groups:

- 43 correspond to postpositions. Some of them can be directly mapped to English prepositions, but in many cases several Basque postpositions correspond to just one English preposition (see Table 1a.). This set also contains postpositions

that map to categories other than English prepositions, such as adverbs (Table 1b).

Table 1. Correspondence between English prepositions and Basque postpositions.

	English	Basque
a.	to	dative (suffix) ablative (suffix) final ablative (suffix)
b.	like	-en gisa (suffix) gisa bezala legez

- 22 types of sentential complements (For instance, English *that* complementizer corresponds to several subordination suffixes: *-la, -n, -na, -nik*).

This shows to which extent the range of arguments is fine grained, in contrast to other works where the range is at the categorial level, such as NP or PP (M. Brent 1993, C. Manning 1993, P. Merlo & M. Leybold 2001).

Due to the complexity carried by having such a high number of cases, we decided to gather postpositions that are semantically equivalent or almost equivalent (for example, English *between* and *among*). Even if there are some semantic differences between them they do not seem to be relevant at the syntactic level. Some linguists were in charge of completing this grouping task. Even considering the risk of making mistakes when grouping the cases, we concluded that the loss of accuracy due to having too sparse data (consequence of having many cases) would be worse than the noise introduced by any mistake in the grouping. The resulting set contained 48 cases. The complexity is reduced but it is still considerable.

Most of the work on automatic acquisition of subcategorization information (J. Carroll & T. Briscoe 1997, A. Sarkar & D. Zeman 2000, A. Korhonen 2001) apply statistical methods (hypothesis testing). Basically the idea is the following: they get "possible subcategorization frames" from automatically parsed data (either completely or partially parsed) or from a syntactically annotated corpus. Afterwards a statistical filter is employed to decide whether those "possible frames" are or not real subcategorization frames. These statistical

methods can be problematic mostly because they perform badly on sparse data. In order to avoid as much as possible data sparseness, we decided to design a system that learns which are the arguments of a given verb instead of learning whole frames. Frames are combinations of arguments, and considering that our system deals with 48 cases, the number of combinations was high, resulting in sparse data. So we decided to work at the level of the argument/adjunct distinction. Working on this distinction is also very useful to avoid noise in the subcategorization frame, because in this task adjuncts are synonyms of noise. A system that tries to get subcategorization frames without previously making the argument/adjunct distinction suffers of having sparse and noisy data.

To accomplish the argument/adjunct distinction we applied two measures: Mutual Information (MI), and Fisher's Exact Test (for more information on these measures, see C. Manning & H. Schütze 1999). MI is a measure coming from Information Theory, defined as the logarithm of the ratio between the probability of the co-occurrence of the verb and the case, and the probability of the verb and the case appearing together calculated from their independent probability. So higher Mutual Information values correspond to higher associated verb and cases (see table 2).

Table 2. Examples from MI values for verb-case pairs

verb	case	MI
atera(to take/go out)	ablative(from)	1.830
atera(to take/go out)	instrumental(with)	-0.955
erabili(to use)	gisa(as)	2.255
erabili(to use)	instrumental(with)	-0.783

Mutual Information shows higher values for *atera-ablative*(to go/take out), *erabili-gisa* (to use-as). These pairs were manually tagged as arguments, therefore Mutual information makes the right prediction. On the contrary, *atera-instrumental* (to go/take out-with), *erabili-instrumental* (to use-with) were manually tagged as adjuncts. Mutual information values in table 2 go along with the manual tagging for these last pairs as well, because the Mutual information values are low as should correspond to adjuncts.

Fisher's Exact Test is a hypothesis testing statistical measure¹. We used the left-side version of the test (see T. Pederssen, 1996). Under this version the test tells us how likely it would be to perform the same experiment again and be less accurate. That is to say, if you were repeating the experiment and there were no relation between the verb and the case, you would have a big probability of finding a lower co-occurrence frequency than the one you observed in your experiment. So higher left-side Fisher values tell us that there is a correlation between the verb and the case (see table 3.)

Table 3. Examples of Fisher's Exact Test values for verb-case pairs

verb	Case	Fisher
atera(to take/go out)	Ablative(from)	1.0000
atera(to take/go out)	instrumental(with)	0.0003
erabili(to use)	gisa(as)	1.0000
erabili(to use)	instrumental(with)	0.0002

Fisher's Exact values show higher values for atera-ablative(to go/take out), erabili-gisa (to use-as). These values predict correctly the association between the verbs and cases for these examples. The low values for the atera-instrumental (to go/take out-with), and erabili-instrumental (to use-with) pairs, should be interpreted as the non-association between the verbs and the cases in these examples, that is to say, they are adjuncts. And again, the prediction would be right according to the taggers.

These tests are broadly used to discover associations between words, but they show different behaviour depending on the nature of the data. We did not want to make any a priori decision on the measure employed. On the contrary, we aimed to check which test behaved better on our data.

3 Evaluation

We found in the literature two main approaches to evaluate a system like the one proposed in this paper (T. Briscoe & J. Carroll 1997, A. Sarkar & D. Zeman 2000, A. Korhonen 2001):

- Comparing the obtained information with a gold standard.
- Calculating the coverage of the obtained information on a corpus. This can give an estimate of how well the information obtained could help a parser on that corpus.

Under the former approach a further distinction emerges: using a dictionary as a gold standard, or performing manual evaluation, where some linguists extract the subcategorization frames appearing in a corpus and comparing them with the set of subcategorization frames obtained automatically.

We decided to evaluate the system both ways, that is to say, using a gold standard and calculating the coverage over a corpus. The intention was to determine, all things being equal, the impact of doing it one way or the other.

3.1 Evaluation 1: comparison of the results with a gold standard

From the 640 analyzed verbs, we selected 10 for evaluation. For each of these verbs we extracted from the corpus the list of all their dependents. The list was a set of bare verb-case pairs, that is, no context was involved and, therefore, as the sense of the given verb could not be derived, different senses of the verb were taken into account. We provided 4 human annotators/taggers with this list and they marked each dependent as either argument or adjunct. The taggers accomplished the task three times. Once, with the simple guideline of the implicational test and obligatoriness test, but with no further consensus. The inter-tagger agreement was low (57%). The taggers gathered and realized that the problem came mostly from semantics. While some taggers tagged the verb-case pairs assuming a concrete semantic domain the others took into account a wider range of senses (moreover, in some cases the senses did not even match). So the tagging was repeated when all of them considered the same semantics to the different verbs. The inter-tagger agreement raised up to a 80%. The taggers gathered again to discuss, deciding over the non clear pairs.

The list obtained from merging² the 4 lists in one is taken to be our gold standard. Notice that

¹ There are two ways of interpreting Fisher's test, as one or two sided test. In the one sided fashion there is still another interpretation, as a right or left sided test.

² Merging was possible once the annotators agreed on the marking of each element.

when the annotators decided whether a possible argument was really an argument or not, no context was involved. In other words, they were deciding over bare pairs of verbs and cases. Therefore different senses of the verb were considered because there was no way to disambiguate the specific meaning of the verb. So the evaluation is an approximation of how well would the system perform over any corpus. Table 4 shows the results in terms of Precision and Recall.

Table 4. Results of Evaluation 1 (context independent)

	Precision	Recall	F-score
MI	62%	50%	55%
Fisher	64%	44%	52%

3.2 Evaluation 2: Calculation of the coverage on a corpus

The initial corpus was divided in two parts, one for training the system and another one for evaluating it. From the fraction reserved for evaluation we extracted 200 sentences corresponding to the same 10 verbs used in the "gold standard" based evaluation. In this case, the task carried out by the annotators consisted in extracting, for each of the 200 sentences, the elements (arguments/adjuncts) linked to the corresponding verb. Each element was marked as argument or adjunct. Note that in this case the annotation takes place inside the context of the sentence. In other words, the verb shows precise semantics.

We performed a simple evaluation on the sentences (see table 5), calculating precision and recall over each argument marked by the annotators³. For example, if a verb appeared in a sentence with two arguments and the statistical filters were recognizing them as arguments, both precision and recall would be 100%. If, on the contrary, only one was found, then precision would be 100%, and recall 50%.

Table 5. Results of Evaluation 2 (inside context)

	Precision	Recall	F-score
MI	93%	97%	95%
Fisher	93%	93%	93%

³ The inter-tagger agreement in this case was of 97%.

3.3 Discussion

It is obvious that the results attained in the first evaluation are different than those in the second one. The origin of this difference comes mostly, on one hand, from semantics and, on the other hand, from the nature of statistics:

- Semantic source. The former evaluation was not contextualized, while the latter used the sentence context. Our experience showed us that broader semantics (non-contextualized evaluation) leads to a situation where the number of arguments increases with respect to narrower (contextualized evaluation) semantics. This happens because in many cases different senses of the same verb require different arguments. So when the meaning of the verb is not specified, different meanings have to be taken into account and, therefore, the task becomes more difficult.
- Statistical reason. The disagreement in the results comes from the nature of the statistics themselves. Any statistical measure performs better on the most frequent cases than on the less frequent ones. In the first experiment all possible arguments are evaluated, including the less frequent ones, whereas in the second experiment only the possible arguments found in the piece of corpus used were evaluated. In most of the cases, the possible arguments found were the most frequent ones.

At this point it is important to note that the system deals with non-structural cases. In Basque there are three structural cases (ergative, absolutive and dative) which are special because, when they appear, they are always arguments. They correspond to the subject, direct object and indirect object functions. These cases are not very conflictive about argumenthood, mainly because in Basque the auxiliary bears information about their appearance in the sentence. So they are easily recognized and linked to the corresponding verb. That is the reason for not including them in this work. Precision and recall would improve considerably if they were included because they are the most frequent cases (as statistics perform well over frequent data), and also because the shallow parser links them correctly using the information carried by the auxiliary. Notice that we did not incorporate them because in the future we would like to use the subcategorization

Table 3. Summary of several systems on subcategorization information.

Method	Number of frames	Number of verbs	Linguistic resources	F-Score (evaluation based on a gold standard)	Coverage on a corpus
C. Manning (1993)	19	200	POS tagger + simple finite state parser	58	
T. Briscoe & J. Carroll (1997)	161	14	Full parser	55	
A. Sarkar & D. Zeman (2000)	137	914	Annotated treebank	-	88
D. Kawahara et al. (2001)	-	23,497	Full parser		82 accuracy
M. Maragoudakis et al. (2001)	-	47	Simple phrase chunker	77	
This paper	-	640	Morph. Analyzer + Phrase Chunker + Finite State Parser	55	95

information obtained for helping our parser, and the non-structural cases are the most problematic ones.

4 Related work

Concerning the acquisition of verb subcategorization information, there are proposals ranging from manual examination of corpora (R. Grishman et al. 1994) to fully automatic approaches.

Table 3, partially borrowed from A. Korhonen (2001), summarizes several systems on subcategorization frame acquisition.

C. Manning (1993) presents the acquisition of subcategorization frames from unlabelled text corpora. He uses a stochastic tagger and a finite state parser to obtain instances of verbs with their adjacent elements (either arguments or adjuncts), and then a statistical filtering phase produces subcategorization frames (from a set of previously defined 19 frames) for each verb.

T. Briscoe and J. Carroll (1997) describe a grammar based experiment for the extraction of subcategorization frames with their associated relative frequencies, obtaining 76.6% precision and 43.4% recall. Regarding evaluation, they use the ANLT and COMLEX Syntax dictionaries as gold standard. They also performed evaluation of coverage over a corpus. For our work, we could

not make use of any previous information on subcategorization, because there is nothing like a subcategorization dictionary for Basque.

A. Sarkar and D. Zeman (2000) report results on the automatic acquisition of subcategorization frames for verbs in Czech, a free word order language. The input to the system is a set of manually annotated sentences from a treebank, where each verb is linked with its dependents (without distinguishing arguments and adjuncts). The task consists in iteratively eliminating elements from the possible frames with the aim of removing adjuncts. For evaluation, they give an estimate of how many of the obtained frames appear in a set of 500 sentences where dependents were annotated manually, showing an improvement from a baseline of 57% (all elements are adjuncts) to 88%.

Comparing this approach to our work, we must point out that Sarkar and Zeman's data does not come from raw corpus, and thus they do not deal with the problem of noise coming from the parsing phase. Their main limitation comes by relying on a treebank, which is an expensive resource.

D. Kawahara et al. (2001) use a full syntactic parser to obtain a case frame dictionary for Japanese, where arguments are distinguished by their syntactic case, including their headword (selectional restrictions). The resulting case frame components are selected by a frequency threshold.

M. Maragoudakis et al. (2001) apply a morphological analyzer and phrase chunking module to acquire subcategorization frames for Modern Greek. In contrast to this work, they use different machine learning techniques. They claim that Bayesian Belief Networks are the best learning technique.

P. Merlo and M. Leybold (2001) present learning experiments for automatic distinction of arguments and adjuncts, applied to the case of prepositional phrases attached to a verb. She uses decision trees tested on a set of 400 verb instances with a single PP, reaching an accuracy of 86.5% over a baseline of 74%.

Note that both Manning and Merlo and Leybold's systems learn from contexts with just one PP (maximum) per verb (finite state filter). Our system learns from contexts with up to 5 PPs. Furthermore, we distinguish 48 different kinds of cases, hence the number of combinations is considerably bigger.

Regarding the parsing phase, the systems presented so far are heterogeneous. While Manning, Merlo and Leybold and Maragoudakis et al. use very simple parsing techniques, Briscoe and Carroll and Kawahara et al. use sophisticated parsers. Our system can be placed between these two approaches. The result of the shallow parsing is not simple in that it relies on a robust morphological analysis and disambiguation. Remember that Basque is an agglutinative language with strong morphology and, therefore, this stage is particularly relevant. Moreover, the finite state filter we used for parsing is very sophisticated (L. Karttunen et al. 1997, I. Aldezabal et al. 2001), compared to Manning's.

Conclusion

This work describes an initial effort to obtain subcategorization information for Basque. To successfully perform this task we had to go deeper than mere syntactic categories (NP, PP, ...) enriching the set of possible arguments to 48 different classes. This leads to quite sparse data.

Together with sparseness, another problem common to every subcategorization acquisition system is that of noise, coming from adjuncts and incorrectly parsed elements. For that reason, we defined subcategorization acquisition in terms of distinguishing between arguments and adjuncts.

The system presented was applied to a newspaper corpus. Subcategorization acquisition is highly associated to semantics in that different senses of a verb will most of the times show different subcategorization information. Thus, the task of learning subcategorization information is influenced by the corpus. As for the evaluation of this work, we carried out two different kinds of evaluation. This way, we verified the relevance of semantics in this kind of task.

For the future, we plan to incorporate the information resulting from this work in our parsing system. We hope that this will lead to better results in parsing. Consequently, we would get better subcategorization information, in a bootstrapping cycle. We also plan to improve the results by using semantic information as proposed in A. Korhonen (2001).

Acknowledgements

This work has been supported by the Department of Economy of the Government of Gipuzkoa, The University of the Basque Country, the Department of Education of the Basque Government and the Commission of Science and Technology of the Spanish Government.

References

- I. Aduriz, J. M. Arriola, X. Artola, A. Díaz de Ilarraza, K. Gojenola and M. Maritxalar (1997) *Morphosyntactic disambiguation for Basque based on the Constraint Grammar Formalism*. Conference on Recent Advances in Natural Language Processing (RANLP).
- I. Alegria, X. Artola, K. Sarasola and M. Urkia (1996) *Automatic morphological analysis of Basque*. Literary and Linguistic Computing. 11 (4), Oxford University.
- I. Aldezabal, K. Gojenola and K. Sarasola (2000) *A Bootstrapping Approach to Parser Development*. International Workshop on Parsing Technologies (IWPT), Trento.
- I. Aldezabal, M. Aranzabe, A. Atutxa, K. Gojenola, M. Oronoz M. and Sarasola K. (2001) *Application of finite-state transducers to the acquisition of verb subcategorization information*. Finite State Methods in Natural Language Processing, ESSLLI Workshop, Helsinki.
- M. R. Brent (1993) *From Grammar to Lexicon: Unsupervised Learning of Lexical Syntax*. Computational Linguistics, 19:243-262.

- T. Briscoe and J. Carroll (1997) *Automatic Extraction of Subcategorization from Corpora*. ANLP-97:356-363.
- J. Carroll, G. Minnen and T. Briscoe (1998) *Can Subcategorization Probabilities Help a Statistical Parser?* Proceedings of the 6th ACL/SIGDAT Workshop on Very Large Corpora, Montreal.
- N. Chomsky (1995) *The Minimalist Program*. Cambridge MA, MIT Press.
- T. Dunning (1993) *Accurate Methods for the Statistics of Surprise and Coincidence*. Computational Linguistics 19, 1
- J.M. Gawron (1986) *Situations and prepositions*. Linguistics and Philosophy 9(3), 327-382.
- J. Grimshaw (1990) *Argument Structure*. Cambridge, MA, MIT Press.
- R. Grishman, C. Macleod, A. Meyers (1994) *Complex Syntax: Building a Computational Lexicon*. COLING-94.
- F. Karlsson, A. Voutilainen, J. Heikkilä, A. Anttila (1995) *Constraint Grammar: A Language-independent System for Parsing Unrestricted Text*. Mouton de Gruyter.
- L. Karttunen, J.P. Chanod, G. Grefenstette, A. Schiller (1997) *Regular Expressions For Language Engineering*. Natural Language Engineering.
- D. Kawahara, N. Kaji and S. Kurohashi (2000) *Japanese Case Structure Analysis by Unsupervised Construction of a Case Frame Dictionary*. COLING-2000, Saarbrücken.
- A. Korhonen (2001) *Subcategorization acquisition*. Unpublished PhD Thesis, University of Cambridge.
- K. Koskenniemi (1983) *Two-level Morphology: A general Computational Model for Word-Form Recognition and Production*. PhD thesis, University of Helsinki.
- J. Kuhn, J. Ecker-Kohler and C. Rohrer (1998) *Lexicon Acquisition with and for Symbolic NLP-Systems -- a Bootstrapping Approach*. First International Conference on Language Resources and Evaluation (LREC98), Granada.
- C.D. Manning (1993) *Automatic Acquisition of a Large Subcategorization Dictionary from Corpora*. Proceedings of the 31th ACL.
- C.D. Manning and H. Schütze (1999) *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Massachusetts.
- M. Maragoudakis, K. Keramidis, N. Fakotakis and G. Kokkinakis (2001) *Learning Automatic Acquisition of Subcategorization Frames using Bayesian Inference and Support Vector Machines*. The 2001 IEEE International Conference on Data Mining, IMDC'01, San José.
- P. Merlo and M. Leybold (2001) *Automatic Distinction of Arguments and Modifiers: the Case of Prepositional Phrases*. EACL-2001, Toulouse.
- T. Pederssen (1996) *Fishing for Exactness* In the Proceeding of the South-Central SAS User Group Conference (SCSUG-96).
- C. Pollard and I. Sag (1987) *An information based Syntax and Semantics*, volume 13. CSLI lecture. Notes, Stanford University.
- A. Sarkar and D. Zeman (2000) *Automatic Extraction of Subcategorization Frames for Czech*. COLING-2000, Saarbrücken.
- C. Schütze (1995) *PP Attachment and Argumenthood*. MIT Working Papers in Linguistics.
- C. Verspoor (1997) *Contextually-Dependent Lexical Semantics*. PhD thesis, Brandeis University, MA.