PARSING THE CHILDES DATABASE: METHODOLOGY AND LESSONS LEARNED

Kenji Sagae* and Alon Lavie* and Brian MacWhinney**

*Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213, USA
{sagae, alavie}@cs.cmu.edu

**Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213, USA
macw@cmu.edu

Abstract

This paper discusses the process of parsing adult utterances directed to a child, in an effort to produce a syntactically annotated corpus of the verbal input to a human language learner. In parsing the Eve corpus of the CHILDES database, we encountered several challenges relating to parser coverage and ambiguity, for which we describe solutions that result in a system capable of analyzing almost 80% of the adult utterances in the corpus correctly. We describe characteristics of the language in the corpus that make this task unique, and present specific ways to deal with the analysis of this type of language. We discuss each step of the corpus analysis in detail, focusing on how selected techniques, such as part-of-speech tagging, rule-based robust parsing and statistical disambiguation, affect the trade-off between coverage and accuracy. Finally, we present a detailed evaluation of the performance of our system. A parsed corpus resulting from the research described in this paper is available to the research community.

1 Introduction

Recent efforts in designing computational natural language learning models have been marked by their extensive use of data (Lieven et al., 1997). Although the availability of large amounts of raw text has been of great importance to such efforts, linguistically annotated corpora have retained a central role in computational natural language acquisition. Syntactically annotated corpora are of particular value in language learning research, providing information not only on the superficial aspects of text, but also on its linguistic structure.

This paper discusses our methodology and the issues we faced in annotating the adult language portion of a corpus of transcribed verbal interactions between a child and her parents with part-of-speech tags and syntactic feature structures. We focus on the use of robust parsing and a number of strategies to manage ambiguity while covering most of the syntactic constructions found in the corpus. We begin by contrasting our approach to other well-known efforts in corpus annotation in section two. In section three we discuss our motivation, details of the corpus we used in our work, and how they relate to our parsing task. We point out specific differences between the child and adult utterances in the corpus, and argue that analyzing those two language types constitute significantly different problems. We then proceed to discuss our approach to the better defined of the two problems: analyzing the adult language in the corpus. Section four describes the general components of our analysis system, and section five explores our strategies to reduce ambiguity and increase coverage in parsing the adult portion of our target corpus. In section six we present the results of our experiments with adult sentences. Finally, in section seven, we discuss the results, present our conclusions, and describe future directions of our work, including the analysis of the child language portion of the corpus.

2 Other Efforts in Annotating Corpora with Syntactic Information

One of the best-known efforts to produce corpora with syntactic annotations is the Penn Treebank project (Marcus, 1993). It currently includes two corpora (one composed of financial news articles, and the other of casual conversational communication) annotated with part-of-speech tags and skeletal syntactic parse trees. The efforts in annotating corpora in the Penn Treebank project included an automated first step (using a part-of-speech tagger and a parser), but relied heavily on the manual efforts of linguists to achieve high-quality linguistic annotations. Although this approach yields a high level of accuracy, it is impractical if time is relatively limited and a team of linguists dedicated to corpus-annotation is not available.

Another well-known related research effort is the Alembic Workbench (Day, 1997), which is an environment for efficient corpus annotation. It offers a graphical user interface and language processing tools to aid the annotation process. The NEGRA Corpus (Skut, 1997) is another example of human-machine collaboration in annotating a corpus with syntactic and grammatical function information. In fact, a number of syntactically annotated corpora (or treebanks) have been produced in recent years, with varying amounts of automation, but typically with human effort playing a major role in the annotation process. Although the study of syntax and grammar can be viewed as the major global concern in all treebanks, some collections address specific concerns in natural language research, from the study of anaphora resolution (McEnery et al., 1997), to parser evaluation (Carrol et al., 1999). For a collection of papers on different treebanks, annotation techniques, methodologies and foci, please see (Garside et al., 1997).

Our work differs in focus from these other projects in that our primary goal is to provide information specific to the process of human language acquisition. Although this information could be used in the training of automatic learning systems, we also aim to make it suitable for researchers in various aspects of linguistics and child development to draw reliable conclusions on the human language learning process. Because of our goals, each annotation should be as error-free as possible. Our methodology involves analyzing the particular kind of spontaneous language parents use when addressing their children, and producing constituent structures and syntactic feature structures that provide further information on that language (such as the assignment of subject, objects and other syntactic roles in a sentence). Our work differs from some of the other projects mentioned in that we focus mainly on the application of robust parsing and parse selection techniques to reduce the manual effort required to annotate a corpus reliably and accurately, placing a strong emphasis on the role of natural language processing tools in corpus annotation.

3 The CHILDES Database and the Eve Corpus

The CHILDES database (MacWhinney, 1995) is a collection of manually transcribed corpora of casual verbal interactions between children and their parents. In addition to transcribed utterances, the child language in some of the corpora in the database is annotated with part-of-speech tags. Because the database contains a large amount of data on human language acquisition, it is of great value for research on human language learning, from a computational as well a psycholinguistic perspective. We believe annotations containing syntactic information would add valuable information to the corpus as data that can be used to study the role and the syntactic content of parental verbal input in the process of human language acquisition. We hope such

```
*CHI:  more cookie.
%mor:  qn|more n|cookie .
```

Figure 1: Sample child utterance from the Eve corpus

```
*MOT:  how about another graham
       cracker ?
%mor:  adv:wh|how
       prep|about^adv|about
       det|another n|graham
       n|cracker ?
```

Figure 2: Sample adult utterance from the Eve corpus

information will be useful in the research of computational models of human language learning, allowing for the development of models strongly rooted on syntactic evidence from data actually used in a child's language acquisition process.

Among the corpora in the CHILDES database, we chose to focus on the Eve corpus (Brown, 1973). Our choice was motivated by the corpus' clean transcription and manually verified part-of-speech tags for the child utterances, as well as its central role in child language acquisition research (Moerk, 1993). The corpus includes utterances from the child (Eve), as well as adults (Eve's parents). An example of child utterances in the corpus can be seen in figure 1. In this example, the first line is a transcription of one of Eve's utterances (indicated by `*CHI:`), and the following line contains part-of-speech and morphology annotations for that utterance (indicated by `%mor:`). Adult sentences in the corpus are also accompanied by a line with part-of-speech information, but that information is produced fully automatically and is often ambiguous. An example can be seen in figure 2.

Our work consists of the use of parsing techniques to analyze each adult utterance in the corpus and produce syntactic annotations in the form of constituent structures (marked by `%cst:`) and syntactic feature structures (marked by `%fst:`), as illustrated in figure 3. The syntactic feature structures produced as annotations resemble those of Lexical Functional Grammar (Bresnan, 2001) in the form of feature-value pairs (although we

```
*MOT:   you kicked it .
%mor:   pro|you v|kick-PAST pro|it .
%fst:   ((mood *declarative) (tense *past)
         (index 2)
         (subject ((cat pro) (num *sg) (pers 2)
                   (case *nom)
                   (index 1) (root *you)))
         (object ((cat pro) (sum sg) (pers 3)
                   (case acc) (index 3)
                   (root *it)))
         (root *kick) (cat v))
%cst:  (sentence (decl (np (pro you))
          (vp (vbar (v kicked)
             (np (pro it)))))
         (period .))
```

Figure 3: Sample syntactic annotations in the Eve corpus

make no attempt to follow LFG theory closely), where typically the features are syntactic functions, and the values are syntactic constituents or syntactic characteristics of the sentence. The `index` features provide a cross-reference between a feature structure and its corresponding constituent structure.

While the adult language in corpora from the CHILDES database is casual and conversational, differing significantly from written natural language, the child language in the corpora varies from the language of a child in the very early stages of language learning, to fairly complex syntactic constructions. We believe that the child and adult utterances differ significantly enough that we may be able to analyze them more accurately by doing so separately, possibly with different strategies. In this paper, we explore the "easier" (in the sense that it is better defined) problem of analyzing the adult utterances in the Eve corpus, whose role in child language acquisition has been the subject of extensive research (Moerk, 1993).

4 The Analysis System

To produce the syntactic analyses necessary for annotation of the corpus, we developed a syntactic analysis system based on robust parsing and statistical disambiguation. The input to our system is a sequence of transcribed utterances, and the output is a syntactic analysis for each of those utterances, as seen in the example in section 3. At a lower level, the system can be divided into three main components:

1. POST, a part-of-speech tagger developed especially for the CHILDES database and its custom set of part-of-speech tags for child-parent communication (Parisse and Le Normand, 2000).
2. LCFlex (Rosé and Lavie, 2001), a robust parser that provides special features for parsing spoken language. Because the corpora in the CHILDES database consist only of transcribed casual conversational speech (with its disfluencies and ungrammaticalities), having a parser designed to handle such language is of great importance. Through a set of parameters, LCFlex can be tuned to allow the insertion of specific missing syntactic constituents into a sentence, and to skip extra-grammatical material that would prevent an analysis from being found with the grammar in use. These features allow great flexibility in parsing spontaneous casual text, but their parameters must be tuned carefully to balance benefits and the increased ambiguity caused by allowing insertions and skipping.
3. A statistical disambiguation module to pick the correct analysis from the many produced by the parser. The idea behind syntactic disambiguation in LCFlex is that each analysis of a particular utterance is obtained through an ordered succession of grammar rule applications, and the correct analysis should be the one resulting from the most probable succession of rules. The probability of each competing analysis is determined based on a statistical model of bigrams of rule applications (Rosé and Lavie, 2001) obtained from training examples consisting of sentences and their correct analyses.

In addition to the automatic components above, a human linguist who provides correctness judgments for part-of-speech tag assignments and syntactic analyses is indispensable in the annotation process. Although some amount of human effort is still needed, one of the goals of our work is to reduce this need to the point where a single person could annotate 80%[1] of a 15,000 sentence corpus in one week.

One of the main challenges faced by an analysis system such as the one just outlined is the trade-off between coverage and ambiguity associated with imperfect analysis models of natural language. One on hand, we must produce an analysis for as many utterances in the corpus as possible, which means that our grammar must cover

[1] This corresponds to the portion of the corpus for which we expect our system to produce correct analyses.

(nearly) all syntactic constructions found in the corpus. On the other hand, allowing the system to have large coverage inevitably results in an increased number of different analyses that can be produced for a single utterance, making it less likely that the correct one would be correctly identified by the system. Because the data produced by the system should be fit for human investigation of a number of different language phenomena within each sentence, where error tolerance may be very low, an analysis is only acceptable if it contains no errors. This simplifies the role of the human linguist to making binary decisions on the correctness of the output provided by the system, promoting overall consistency and efficiency in the annotation process.

5 Tailoring a High Performance Analysis System

Now we discuss our steps in addressing coverage and ambiguity issues while providing the system with the necessary knowledge to analyze the Eve corpus.

5.1 Grammar

The grammar used by LCFlex consists of context-free rules augmented with feature unification constraints. Although general-purpose English grammars are available, we found that they were not suitable for our analysis task. The main problems associated with "off-the-shelf" grammars are related to the large amount of ambiguity allowed by such grammars, and the lack of support for certain phenomena we find in corpora in the CHILDES database, such as the extensive use of communicators[2] and vocatives commonly used in casual spoken language, onomatopoeia, etc. Because of the nature and the domain of our target corpus, it does not contain many of the complex syntactic constructions found in newspaper-style text, or even adult conversations. We can take advantage of this fact, and attempt to reduce ambiguity by tailoring a grammar that fits our target language more tightly.

Starting with a general-purpose English grammar with about 600 rules, we pruned or simplified a large number of rules that would never (or rarely) be used in correct analyses for the target corpus. The result was a completely rewritten compact grammar with 152 rules. This final grammar included rules to handle the specific language phenomena likely to appear in the CHILDES database, and represents a much cleaner and tighter model of the language we are attempting to analyze. As a result, the potential for ambiguity in parsing was significantly reduced.

5.2 Lexical Ambiguity

Even though a more suitable grammar is a first step towards managing ambiguity, it is not a complete solution to the problem, and further techniques to resolve syntactic ambiguity are needed. One such way is to eliminate lexical ambiguity by selecting a single part-of-speech tag for each word, using the part-of-speech tagger. The first step is to have a corpus of correctly tagged text to train the tagger. Unfortunately, the CHILDES database contains no unambiguous part-of-speech tagged data for adult utterances. While tagged data for child utterances are available, the child and adult languages are significantly different so that a tagger trained on child utterances

[2] The *communicator* part-of-speech tag is used in the annotation scheme of the CHILDES database to denote extra-grammatical words used to grab attention verbal interactions, but usually devoid of meaning or syntactic role in a sentence (MacWhinney, 1995).

would perform poorly in tagging adult ones. To create a part-of-speech tagging training corpus for the adult language in the corpus, we used the following bootstrapping process:

1. Use tagged child utterances to train a part-of-speech tagger for adult utterances.
2. Tag adult utterances (4,000 words) and hand correct them.
3. Retrain the tagger with the newly corrected data, and iterate from step 2.

By performing four iterations of the procedure above, we improved the accuracy of the part-of-speech tagger from an initial 87.2% to 94.3%. The improvement in accuracy for each iteration decreased at a rapid pace, and it is unlikely that further iterations would yield significant benefits (at least not cost-effectively).

5.3 Syntactic Ambiguity

Once syntactic ambiguity has been reduced through the elimination of lexical ambiguity, we can attempt to find the single correct analysis produced by the parser (when one exists) using statistical disambiguation. For that, we need a training corpus of correct sentence-analysis pairs. We create this data in a way similar to the bootstrapping process used to generate part-of-speech training data, but this time we start with the results of parsing lexically unambiguous input.

1. Parse part-of-speech tagged utterances.
2. Examine unambiguous (or nearly unambiguous) analyses.
3. Add correct analyses to training corpus.
4. Train statistical disambiguation module.
5. Use iterative process similar to the one used for obtaining the part-of-speech training corpus.

We started with an initial statistical disambiguation training corpus of less than 500 sentences, and grew its size to 3,000 sentences in four iterations of the process described above. As with the process for building a part-of-speech training corpus, the benefits of successive iterations decreased at a fast rate.

Although only a small number of sentences can by fully syntactically disambiguated by eliminating lexical ambiguity alone, those sentences can be very useful in obtaining an initial training corpus for the statistical disambiguation module, since resolving large amounts of syntactic ambiguity manually may be a practically intractable task.

As a way to enhance the performance of the procedures described above to create training corpora for part-of-speech tagging and syntactic disambiguation, we can exploit the interactions between part-of-speech tagging and parsing. Once we obtain an initial statistical disambiguation training corpus, we can increase its size while also increasing the size of our part-of-speech tagging training corpus by using a feedback loop between part-of-speech tagging and parsing. We assume that the input sentences for which the parser produces correct analyses have correct part-of-speech tag assignments, and we add those sentences to our part-of-speech training corpus. Improvements in part-of-speech tagging, in turn, result in more correct analyses being produced by the parser.

5.4 Parser Flexibility

Even after grammar development, a large number of sentences in the Eve corpus still could not be parsed with our compact grammar due to specific characteristics of the casual conversational language in the corpus (not due to general syntactic structures). The majority of such sentences were not covered successfully because of omitted words or filled pauses in otherwise fully grammatical utterances, for example:

- Missing auxiliary verbs in questions ("[*Do*] You want to go outside?");

- Missing noun phrases, as elided subjects ("[*I*] Don't think she wants to play now."), and even as elided objects ("Give [*it*] to me.");
- Missing auxiliary verbs and noun phrases ("[*Do*] [*you*] Want to go outside?");
- Filled pause ("I'd like to tell you, *uh*, something.").

Adding explicit ad-hoc grammar rules to handle such sentences would cause the grammar to deviate from the clean model of language we were hoping to achieve, and add much harmful ambiguity to the analysis process. Instead, we turned to the robustness features of the parser to handle these sentences. LCFlex allows the addition of specific syntactic nodes to an analysis, or skipping of words in a sentence, making it conform to the grammar and leading to a successful analysis.

5.5 Balancing Coverage and Ambiguity

We now examine the effects of each of the strategies above on the coverage/ambiguity trade-off.

Decreasing Ambiguity

Parsing with the initial general English grammar and without proper training of the disambiguation module yields very few correct analyses due to ambiguity. We only consider an analysis "correct" if it contains no errors or ambiguity. Using the same grammar coupled with the statistical disambiguation provided by LCFlex, we obtain less than 50% accuracy in analyzing the Eve corpus (measured with a 200 utterance test corpus). We define accuracy as the ratio between correct analyses and the total number of sentences analyzed. Using our final rewritten grammar and statistical disambiguation (trained on 3,000 correctly parsed utterances), we get close to 65% accuracy. This reflects both the improvement in ambiguity resolution and some gain in coverage.

Using the final grammar and part-of-speech tagged input sentences to eliminate lexical ambiguity, the number of correct analyses decreases to 57.5%. However, the set of correct analyses obtained with this setup is not a subset of the set of correct analyses obtained with lexically ambiguous input. Although the ratio of correct analyses over non-failed analyses increases, the system fails to analyze (and produces no output for) a large number of utterances due to errors in part-of-speech assignments. In terms of the trade-off, we decreased ambiguity significantly, but at the cost of a severe reduction in coverage. We achieved a 1.1% improvement in part-of-speech tagging using transformation-based learning of Brill-style rules (Brill, 1995), which resulted in a slight improvement in coverage. However, overall parser accuracy obtained with part-of-speech tagged input was still under 60%.

Increasing Coverage

Setting the parser to allow limited insertions (a single noun-phrase and/or a single auxiliary may be inserted during parsing) performs well, as expected, for about 5% of the sentences in the Eve corpus. However, the percentage improvement in accuracy is less than 3%, due to the increased ambiguity and over-generation that results from increasing the search space of possible analyses with insertions. Allowing limited skipping (a single word in the input utterance may be skipped during parsing) actually decreases the overall accuracy. In other words, the number of sentences that are parsed incorrectly due to the increased search space is greater than the number of correct analyses that result from limited skipping.

Pass	POS Ambiguity	Insertion	Skipping
1	None	None	None
2	Limited	None	None
3	Limited	Auxiliary	None
4	Limited	NP	None
5	Limited	Auxiliary and NP	None
6	Limited	None	One word

Table 1: Coverage and ambiguity settings for different passes of parsing

Putting it all together

Each of the strategies to reduce ambiguity or increase coverage described above has a different impact on the coverage/ambiguity trade-off, and the effect of applying them together by naively combining them all at once is far from optimal. In summary, our efforts to reduce ambiguity come at the cost of reducing coverage, and our efforts to increase coverage result in much increased ambiguity. By applying lexical disambiguation, limited insertions and skipping, and relying only on the statistical model of bigrams of rule applications for parse selection, we achieve less than 70% parsing accuracy with the Eve corpus.

We must then attempt to balance the coverage/ambiguity trade-off to benefit from both decreased ambiguity and increased coverage. We do so by controlling the amount of ambiguity and coverage in several passes of parsing. We start with the most restrictive settings and the least ambiguity, and upon failures in parsing, gradually increase coverage (and ambiguity). The idea is that we only pay the cost of an increased search space as it becomes necessary, taking advantage of both more limited ambiguity when possible, and increased coverage when needed. Through empirical observation, we arrived at the settings shown in table 1 for each of the passes. In the first pass, we parse lexically unambiguous input, and use no coverage-enhancing techniques. From passes two through six, we allow limited lexical ambiguity, and gradually increase coverage through the robust parsing features of LCFlex. Limited lexical ambiguity means that not every possible part-of-speech tag (according to a lexicon available with the CHILDES database) is allowed for each lexical item, which would cause a greater increase in syntactic ambiguity. Instead, we only allow lexical ambiguity for certain lexical categories where the automatic part-of-speech tagger was observed to make frequent mistakes, causing parser failures. We determined those highly confusable parts-of-speech simply by analyzing the cause of failed analyses, and keeping track of the parts-of-speech most frequently associated with those failures. The following sets of tags accounted for more than 95% of failures caused by a part-of-speech tagging error: {verb, auxiliary}, {verb particle, preposition}, {adverb, adjective}, {noun, verb}.

The reason for not combining multiple coverage-increasing techniques in further passes of parsing is that we prefer having no analysis for an utterance to having an analysis that is very likely to be incorrect. This multi-pass approach not only increases ambiguity gradually only as needed, but also allows us to have some sense of how confident we are that an analysis is correct.

6 Results

6.1 Evaluation

To assess the effectiveness of our methods, we evaluated our current system on 200 randomly chosen previously unused utterances from the Eve corpus, and checked their generated syntactic analyses for correctness. The

Correct Analyses	
Pass 1 (unambiguous POS, no robustness)	115
Pass 2 (ambiguous POS)	29
Pass 3 (insertion of AUX)	3
Pass 4 (insertion of NP)	2
Pass 5 (insertion of AUX and NP)	4
Pass 6 (one word skipping)	4
Total	157 (78.5%)

Table 2: Contribution of each pass to correct analyses

contribution of each of the six passes to the total number of correct analyses can be seen in the table 2. The causes of errors in incorrect analyses are shown in table 3. The row labeled "insertion" refers to the utterances that were not covered by the grammar but were assigned an incorrect analysis due to limited insertions. The row labeled "over-generation" refers to utterances for which the parser did not produce an appropriate analysis due to lack of grammar coverage, but where the utterance was still covered in an incorrect way due to grammar over-generation. Finally, the causes of parsing failures where no analysis was produced for an utterance is shown in table 4.

6.2 Availability

One of the main goals of this project is to provide data to the language acquisition research community. The current results of the research described in this paper (a version of the Eve corpus with syntactic annotations for adult utterances), as well as related tools and other resources, are available for research purposes at the CHILDES web site (http://childes.psy.cmu.edu), or by request from the authors. It is our hope that the data we have produced will be useful in current research efforts in language acquisition[3], as well as inspire and fuel new research on natural language learning and various aspects of grammar acquisition.

7 Conclusions and Future Work

Our system is quite effective in producing accurate syntactic annotations for the adult utterances in the Eve corpus. The number of incorrect analyses is acceptably small, making the task of manually checking and possibly correcting the resulting annotations fairly manageable, or even unnecessary if an error rate of less than 10% can be tolerated. Practically all of the utterances that failed to be analyzed by the system were not handled

Incorrect Analyses		
Lack of grammar coverage	Insertion	7
	Over-generation	5
POS tag error		4
Transcription error		1
Total		17 (8.5%)

Table 3: Causes of errors in incorrect analyses

No Analysis found	
Lack of grammar coverage	19
Lack of knowledge	5
Transcription error	1
Ungrammatical utterance	1
Total	26 (13%)

Table 4: Causes of parsing failures

[3] See the word order acquisition investigation in (Villavicencio 2000), which has used similar data in lesser amounts, for an example.

due to the occurrence of rare syntactic constructions. Although adding grammar coverage to properly handle these constructions would require the addition of grammar rules, such rules are likely to be very specific and thus do not represent an ambiguity hazard.

Although our efforts to produce syntactic annotations for the child utterances in the corpus is still in very early stages, a preliminary evaluation of the current system on a set of such utterances revealed that more than 60% of them could probably be analyzed correctly with the system as-is. However, significant changes to the overall system would be necessary for analyzing a high percentage of the child utterances accurately and reliably. We are currently working on a different analysis strategy for child utterances, which acknowledges both the global (utterance level) differences and local (fragment or constituent level) similarities between the child and adult languages in the corpus. The analyses produced with this strategy report constituents found in child utterances, without trying to combine them into single global structures when the utterances are ill formed (according to our adult grammar). Our initial heuristic in searching for these constituents is to try to cover as much of the utterance as possible, with as few constituents as possible. Although we recognize the simplistic nature of this approach, our preliminary experiments have yielded very promising levels of accuracy in the analysis of child utterances in the Eve corpus. Further research on analyzing child language is planned as the immediate next step in our work. We also plan to investigate of the effectiveness of the current system on other corpora in the CHILDES database, and possibly the automatic adaptation of the system to other corpora.

Another direction we intend to investigate is the improvement of parsing efficiency, making the system more suitable for on-line real-time use. In particular, we plan to use a single parsing pass including lexical disambiguation, limited insertions and skipping, and apply our coverage/ambiguity management techniques as a parse-selection procedure. This would differ from the naïve combination of techniques described in the beginning of section 5.5.3 in that we would not rely solely on statistical disambiguation, but rather on its combination with the heuristics we have described for controlling ambiguity and coverage. Parse selection would be performed based both on a statistical model of rule applications and on the robustness and lexical disambiguation techniques used in each of the possible parses.

Finally, we intend to tune and investigate the effectiveness of our general techniques in other corpora of conversational language in different domains.

8 Acknowledgments

We would like to thank Carolyn Penstein Rosé for her help with LCFlex, Christophe Parisse for his help with POST (in particular the implementation of transformation-based rule learning), Maria Lamendola for her correctness judgments of parser output analyses, and the anonymous reviewers for their insightful comments and suggestions.

9 References

Bresnan, J. (2001). *Lexical-functional syntax*. Blackwell publishers, Oxford, UK.

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

Brown, R. (1973). *A first language: The early stages*. Harvard, Cambridge, MA, USA.

Carrol, J., Minnen, G. and Briscoe, T. (1999) Corpus annotation for parser evaluation, *Journée(s) ATALA sur les corpus annotés pour la syntaxe*. Paris, France.

Day, D., Aberdeen, J., Hirschmann, L., Kozierok, R., Robinson, P., and Vilain, M. (1997). Mixed-initiative development of language processing systems, *Proceedings of the Fifth Conference on Applied Natural Language Processing (ANLP-97)*. Washington, DC, USA.

Garside, R., Leech, G., and McEnery, A. (eds.) (1997) *Corpus Annotation: Linguistic Information from Computer Text Corpora*. Longman, London, England.

Lieven, E., Pine, J., and Baldwin, G. (1997) Lexically-based learning and early grammatical development, *Journal of Child Language*, 24, 187-219.

MacWhinney, B. (1995). *The CHILDES project: Tools for analyzing talk*, Lawrence Erlbaum Associates, Hillsdale, NJ, USA.

McEnery, A.M., Tanaka, I. and Botley, S.P. (1997) Corpus annotation and reference resolution, in Mitkov, R. & Boguraev, B. (eds) *Proceedings of the Association for Computational Linguistics Workshop on Anaphora Resolution for Unrestricted Texts*, Madrid.

Marcus, M. P., Santorini, B., and Marcinkiewics, M. A. (1993). Building a large annotated corpus of English: the Penn Treebank, *Computational Linguistics*, vol 19.

Moerk, E. (1983). *The mother of Eve as a first language teacher*. ABLEX, Norwood, N.J., USA.

Parisse C., and Le Normand, M.T. (2000) Automatic disambiguation of morphosyntax in spoken language corpora, *Behavior Research, Methods, Instruments and Computers*, 32, 3, 468-481.

Rosé, C. P., and Lavie, A. (2001). Balancing robustness and efficiency in unification-augmented context-free parsers for large practical applications, *Robustness in language and speech technology*, van Noord and Junqua (eds.), ELSNET series, Kluwer Academic Press.

Skut, W., Krenn, B., Brants, T, and Uszkoreit, H. (1997). An Annotation Scheme for Free Word Order Languages, *Proceedings of the Fifth Conference on Applied Natural Language Processing (ANLP-97)*. Washington, DC, USA.

Villavicencio, A. (2000). The acquisition of word order by a computational learning system, *Proceedings of the 2nd Learning Language in Logic Workshop*. Lisbon, Portugal.