Proceedings of the 6th Workshop on Very Large Corpora (WVLC-98), 134-142.

What grammars tell us about corpora: the case of reduced relative clauses

Paola Merlo
LATL-University of Geneva
IRCS
University of Pennsylvania
3401 Walnut St Suite 400A
Philadelphia PA 19104-6228
U.S.A.
merlo@linc.cis.upenn.edu

Suzanne Stevenson
Dept of Computer Science
and Center for Cognitive Science (RuCCS)
Rutgers University
CoRE Building, Busch Campus
New Brunswick, NJ 08903
U.S.A.
suzanne@ruccs.rutgers.edu

Abstract

We present a large (65 million words of Wall Street Journal) and in-depth corpus study of a particular syntactic ambiguity to investigate (1) to what extent the structure of a grammar is reflected in a corpus, and (2) how probability functions defined according to a grammar fit independently established measures of syntactic disambiguation preference. We look at the well-known case of the ambiguity between a main clause and reduced relative construction. We measure the probability distributions of several linguistic features (transitivity, tense, voice) over a sample of optionally intransitive verbs. In agreement with recent results on parsing with lexicalised probabilistic grammars (Collins, 1997; Srinivas, 1997), we find that statistics over lexical, as opposed to structural, features best correspond to human intuitive judgments and to experimental findings. These results are enlightening to investigate novel uses of corpora, by assessing the portability of statistics across tasks, and by determining what is needed for useful syntactic annotation of corpora.

1 Introduction

Most linguistic work until the 1950s studied language use, which required attention to detail and exceptions, and led to the development of data-driven theories and to the use of corpora to model naturally occurring language. Later on, linguists mostly studied grammars, which focussed on generalities and regularities, and led to the formulation of strong theories and to the study of similarity across languages. Some of the current "empirical" approaches integrate the corpus-based lessons with the depth of insight that the study of grammar has brought to the study of language.

Empirically-induced models that learn a linguistically meaningful grammar (Collins, 1997) seem to give the best practical results in statistical natural language processing. One of the reasons why these models perform so well compared to probabilistic context-free grammars is that they incorporate detailed lexical knowledge at all points in the derivation (Charniak, 1997). At the same time they perform better than string-based approaches because they retain structural knowledge, such as phrase structure, subcategorization and long distance dependencies. So they are equally capable of modelling the fine lexical idiosyncrasies and the more general syntactic regularities.

Given an annotated training corpus, such methods learn its distributions (the lexical cooccurrences), which requires being given the correct space of events in the model—that is, the grammar—accurately enough that they can parse new instances of the same corpus. The success of such models suggests that a statistical model must have access to the appropriate linguistic features to make accurate predictions.

We might want to ask the question: what happens if what one wants to do with annotated text is *not* to annotate more text, but to perform some other task? Are the same insights valid, so that annotated text can be used to help in other tasks, for instance generation or translation? Can we use annotated text to investigate properties of language(s) systematically? In other words, can we use annotated text as a repository of information? The answer is a qualified yes.

In this paper we look at one type of information that is plentifully present in a corpus—syntactic preferences—and we argue that corpora can be very useful even for tasks that do not involve parsing directly, but that making corpora useful for other tasks might require more a priori information than expected. Precisely, we ask the following question: are the percentages of occurrence of linguistically-defined units in a large corpus in accord with what is known about preferences for these units collected in other ways, such as unedited sentence production, experimental findings, or intuitive native speakers' judgments?

This question is relevant as there is evidence in the literature of human parsing preferences that is in apparent disagreement with predictions of preferences derived from frequencies in a corpus (Brysbaert et al., 1998). Beside the interest in modelling human performance (which is, however, not the focus of the current paper), it is important to investigate the sources of this disagreement between production preference data (frequencies in a text) and perception data (parsing preferences by humans), if the plentiful information stored in text is to be used successfully. Distributional properties of text, if understood, can be used to approximate resolution of ambiguity in several tasks which involve deeper natural language understanding: a generation system can use distributional properties to reproduce users' preference data; automatic translation can use monolingual distributions to model cross-linguistic variation accurately, and automatic lexical acquisition can use distributional properties of text to bootstrap a process of organisation of lexical information.

The method we use to address the question is as follows. We present a large in-depth corpus-based case study (65 million words of WSJ) to investigate (1) how the structure of a grammar is reflected in a corpus, and (2) how probability functions defined according to a grammar fit native speakers' linguistic behaviour in syntactic disambiguation. We look at the well-known case of the ambiguity between a main clause and a reduced relative construction, which arises because regular verbs in English present an ambiguity between the simple past and the past particle (the -ed form). We measure the probability distributions of several linguistic features (transitivity, tense, voice) over a sample of optionally intransitive verbs. We do this by hypothesizing and testing several probability functions over the sample. In agreement with recent results on parsing with lexicalised probabilistic grammars (Collins, 1997; Srinivas, 1997; Charniak, 1997), our main result is that statistics over lexical features best correspond to independently established human intuitive preferences and experimental findings.

We discuss several consequences. Methodologically, this result casts light on the relationship between different ways of collecting preference information. It shows that some apparently contradictory results that have been discussed in the literature can be reconciled. The crucial factor is the level of specificity one looks at. Theoretically, not all lexical features are equally good predictors of linguistic behaviour, and they vary in their ability to correctly classify linguistic phenomena. Finally, from the point of view of language engineering, this results provides a strong indication on what units might port better across tasks, and what are the features that would be most useful in a syntactically annotated corpus..

2 Reduced Relative Clauses

2.1 Linguistic Properties

The following classic "garden-path" example demonstrates the severe processing difficulty that can be associated with the main verb/reduced relative (MV/RR) ambiguity (Bever, 1970):

(1) The horse raced past the barn fell.

Problems arise here because the verb *raced* can be interpreted as either a past tense main verb, or as a past participle within a reduced relative clause (i.e., *the horse* [that was] *raced past the barn*). Because *fell* is the main verb of (1), the reduced relative interpretation of *raced* is required for a coherent analysis of the complete sentence. But the main verb interpretation of *raced* is so strongly preferred that the human language processor breaks down at the verb *fell*, unable to integrate it with the interpretation that has been developed to that point.

This construction is representative of the problem we want to address. It is very *frequent* (MacDonald et al., 1994), hence it constitutes a problem that is relevant for any application. It is both lexically and structurally ambiguous, so it constitutes a *hard* problem. It is well-studied: there are plentiful data on human processing and their relation to frequency of the stimuli (MacDonald, 1994; Trueswell, 1996; Trueswell

Verb Type	example	Judgment
unergative	The horse raced past the barn fell	hard
unaccusative	The butter melted in the pan was rancid	easy
object-drop	The player kicked in the soccer game was angry	easy

Table 1: Processing difficulty of different classes of optionally intransitive verbs according to speakers' intuitions

et al., 1994).

Over the last several years, it has become clear that not all reduced relatives are as difficult as sentence (1) above, and that the difficulty in processing reduced relatives is directly linked to the lexical items in the sentence. In particular the difficulty appears to be related to the type of verb which is involved in the ambiguity. For the ambiguity to arise, the verb involved—*raced* in this case—must be optionally transitive. English has three types of optionally transitive verbs, which differ both in their lexical semantics and in their syntactic properties.

Sentence (1) uses a manner of motion verb, *raced*. In English, these verbs form a subclass of unergative verbs (Levin and Rappaport Hovav, 1995), intransitive action verbs that may appear in a transitive form:

(2a) The horse raced past the barn.

(2b) The rider raced the horse past the barn.

The transitive form of an unergative (2b) is the causative counterpart of the intransitive form (2a), in which the subject of the intransitive becomes the object of the transitive (Hale and Keyser, 1993; Levin and Rappaport Hovav, 1995). Sentences (3a) and (3b) use an unaccusative verb, *melt*:

(3a) The butter melted in the pan.

(3b) The cook melted the butter in the pan.

Unaccusatives are intransitive change of state verbs which also have a causative transitive form. They differ from unergatives because their alternating theta role is a theme (*butter*), while for unergatives it is an agent (*horse*). Finally, sentences (4a) and (4b) use an object-drop verb, *kicked*; these verbs have a non-causative transitive/intransitive alternation, in which the object NP is optional:

(4a) The player kicked the referee.

(4b) The player kicked.

2.2 Processing Difficulty

(Stevenson and Merlo, 1997) asked naive informants for acceptability judgments on sentences with reduced relatives (RRs) containing these verbs. They found that unergative verbs, such as *raced* or *jumped*, uniformly led to a severe garden path in the RR construction, while unaccusative verbs were overwhelmingly judged completely fine in the RR, with a few responses of them being slightly degraded They did not ask for judgments on object-drop verbs; native speakers' intuitions are that they are readily interpretable in a RR. Support for this view comes from experiments which included object-drop verbs, that showed that RRs are relatively easy to understand given a context that is not strongly biased toward a main verb reading (MacDonald, 1994). Thus, the difficulty of the RR interpretation patterns along verb class lines, with unergatives difficult, and unaccusatives and object-drop verbs relatively easy. We summarise these results in Table 1.

2.3 Statistical Properties

We measured the probability distributions of several linguistic features (transitivity, tense, voice) over a sample of optionally transitive verbs from the three lexical semantic classes described above. We proceeded by hypothesizing and testing several probability functions over the sample, and proposing an event classification that best fits the native speaker judgments described above.

In our view, a grammar is a way of classifying elements in a language. Our sample of language is a text, and our grammar is the space of elementary events we define on the text. So our grammar is the space of events over which we calculate the probability distributions. The emphasis on lexicalised grammars, both in linguistics, sentence processing and statistical NLP, points towards statistics computed at the level

	RR	MV	Pass	Act	Trans	Intr	PRT	MV	ADJ	nonADJ
Unergatives	7	4910	139	5330	463	5065	647	4910	21	626
Unaccusatives	21	3321	717	3930	2402	2359	1476	3321	155	1321
Object-drops	202	2316	1339	3074	3355	922	1939	2316	176	1719

Table 2: Raw Counts

of lexical items or their subfeatures.

A *probability space* is a triple Ω, $\mathcal{F}$, P, where Ω is the sample space, $\mathcal{F}$ is the event space and P is a function $P: F \rightarrow [0,1]$. In the discussion below, we assume 5 different probability spaces, in which the event space is defined by sublexical properties of verbs.

First, we counted the occurrences of the verbs as a simple past main verb (MV) and the occurrences of the verbs as a reduced relative (RR). Second, we counted the occurrences of the verbs in a transitive (TRANS) or intransitive (INTR) form. Third, we counted the occurrences of the verbs in an active (ACT) or passive (PASS) form. Then, we counted the occurrences of the verbs as a simple past main verb (MV) and the occurrences of the verbs as a past participle (PRT). These features were chosen because they minimally distinguish main clause from reduce relative forms. Finally, we counted how often the past participle form was used adjectivally. This last count was chosen because only certain lexical semantic classes of verbs (excluding unergative verbs) can occur as adjectives (Levin and Rappaport 1986).

Precisely, $\mathcal{F}'= \{MV, RR\}$, $\mathcal{F}''= \{TRANS, INTR\}$, $\mathcal{F}'''=\{PASS, ACT\}$, $\mathcal{F}''''=\{MV, PRT\}$, $\mathcal{F}'''''= \{NON-ADJ, ADJ\}$. In all cases, we assume that the probabilities of the events are indicated by their relative frequency.

We test the following hypothesis:

> H_0: differences in processing preferences correspond to differences in the distributions of the measured variables.

2.3.1 Materials and Method

We chose a set of 10 verbs from each class, based primarily on the classification of verbs in (Levin, 1993): the unergatives are manner of motion verbs (*jumped*, *rushed*, *marched*, *leaped*, *floated*, *raced*, *hurried*, *wandered*, *vaulted*, *paraded*), the unaccusatives are verbs of change of state (*opened*, *exploded*, *flooded*, *dissolved*, *cracked*, *hardened*, *boiled*, *melted*, *fractured*, *solidified*), and the object-drop verbs are unspecified object alternation verbs (*played*, *painted*, *kicked*, *carved*, *reaped*, *washed*, *danced*, *yelled*, *typed*, *knitted*). Each verb presented the same form in the simple past and in the past participle, as in the MV/RR ambiguity. All verbs can occur in the transitive and in the passive. The verbs in the three sets were matched pairwise in frequency, and their logarithmic frequency varies between 2 and 4 inclusive.

In performing this kind of corpus analysis, one has to take into account the fact that current corpus annotations do not distinguish verb senses. The verbs in the materials were chosen because they did not show massive departures from the intended verb sense; for example, in a different study *run* was eliminated because it occurs most often in phrases such as *run a meeting*, where it is not a manner of motion use. However, in these counts, we did not distinguish a core sense of the verb from an extended use of the verb. So, for instance, the sentence *Consumer spending jumped 1.7 % in February after a sharp drop the month before* (from Wall Street Journal 1987) is counted as an occurrence of the manner-of-motion verb *jump* in its intransitive form. This is an assumption that is likely to introduce more variance than if we had only counted core senses of these verbs, but it is an unavoidable limitation at the current state of annotation of corpora.

Counts were performed on the tagged version of the Brown Corpus and on the portion of the Wall Street Journal distributed by the ACL/DCI (years 1987, 1988, 1989), a combined corpus in excess of 65 million words. Five pairs of counts were collected, for which the raw aggregated results are shown in Table 2. First, each verb was counted in its main verb (i.e., simple past) and past participle uses, based on

Property	Verb Type	Results	
main verb	unerg/unacc	F(18)= 4.058	p=0.059
	unacc/obj-drop	F(18)= 1.498	p=0.237
simple past	unerg/unacc	F(18)=14.927	p=0.001
	unacc/obj-drop	F(18)= 0.317	p=0.580
active	unerg/unacc	F(18)= 9.578	p=0.006
	unacc/obj-drop	F(18)= 0.067	p=0.799
intransitive	unerg/unacc	F(18)= 7.487	p=0.014
	unacc/obj-drop	F(18)= 2.514	p=0.130
non-adj	unerg/unacc	F(14)=13.283	p=0.003
	unacc/obj-drop	F(14)= 0.311	p=0.586

Table 3: Results of anovas

the part of speech tag of the verb in the corpora. Second, active and passive uses of the verbs were counted; cases in which usage could not be determined by a simple pattern search were classified by hand. The third count also required manual intervention: verbs were initially classified as transitive or intransitive according to a set of regular search patterns, then individual inspection of verbs was carried out to correct item-specific errors. In the fourth count, uses of the verb form as main verb or as reduced relative were collected. Reduced relatives were counted by hand after extracting from the corpus all occurrences of the past participle preceded by a noun. In the fifth count, uses of the verbs as prenominal adjectives were counted. None of the verb forms are explicitly marked as adjectives in these corpora. To determine the counts of adjectival uses, we simply divided the verb occurrences labelled with the past participle part of speech tag into prenominal and other uses. The only unexpected result we found was the occurrence of unergative adjectival forms. On inspection all these forms occurred with two verbs: *hurried* occurred 20 times, and *rushed* once. These were not the causative use of the verb. So these verbs were removed from the analysis of variance reported below. The unaccusatives and object-drops that were matched in frequency to *hurried* and *rushed* were also removed (unaccusatives: *boiled*, *fractured*; object-drop: *danced*, *typed*).

2.4 Results and Discussion

The raw aggregated data in Table 2 show that properties related to the main verb (MV) usage—intransitivity, active voice, non-adjectival use and simple past use, as well as the MV construction itself—were more frequent for unergatives than for unaccusatives, and more frequent for unaccusatives than for object-drop verbs. The numerical trend is in accord with the simplest explanation on the use of frequency by humans: more frequently occurring structures are preferred over less frequent alternatives. However, not all numerical differences are significant, as indicated in Table 3.

The data in Table 2 were entered in 10 different analyses of variance on the proportion of cases that indicate a use of the verb as a main verb and its related lexical and sublexical properties — simple past, active, intransitive and non-adjectival use. Results of the ANOVAs are shown in Table 3. The ANOVAS were run to determine if verbs that belong to a class have a significantly different distribution than verbs that belong to one of the other two classes. We chose to perform analysis of variance because this test compares variance within a group to variance between groups, thus it is not distorted by the fact that there is great variation from lexical item to lexical item within each group.

A simplified summary of the results and the corresponding human intuitive data is given in Table 4.

All the data sets show the same pattern. For the lexical features— simple past, active, intransitive and non-adjectival use — the differences between the unergative and unaccusative distributions for each property are highly significant ($p < 0.05$), but the differences between

VERB TYPE	JUDGMENT	SIGNIFICANCE TEST				
		MV/RR	MV/PTR	ACT/PASS	INTR/TRANS	ADJ/NON-ADJ
unergative unaccusative	hard easy	non-sig	sig	sig	sig	sig
unaccusative object-drop	easy easy	non-sig	non-sig	non-sig	non-sig	non-sig

Table 4: Processing difficulty of different classes of optionally intransitive verbs according to speakers' intuitions compared to the results of significance test on pairwise comparisons of corpus data

the unaccusative and object-drop distributions are not ($p > 0.05$). This could explain why the unergatives are significantly more difficult in the RR, while the other classes of verbs are not perceived as different.

Interestingly, a more direct count of the construction itself (the MV/RR probability space) gives different results. Numerically, the counts of RR for unaccusatives are very small, but native speakers do not find RR with unaccusative verbs particularly difficult. Statistically, unergatives are not significantly different from unaccusatives ($p = 0.059$), but native speakers find RR with unaccusative verbs considerably easier than with unergatives.

The picture that emerges from these findings is coherent and in accordance with current developments in statistical parsing and grammatical theory in two important respects. First, the discrepancy between the frequencies of each of the lexical features and the frequencies of the actual construction suggests that the frequency of a construction is a composition function of (at least some of) its lexical features, even if such features are not-independent. Models that can handle non-independent lexical features have given very good results both for part-of-speech and structural disambiguation (Ratnaparkhi, 1996; Ratnaparkhi, 1997; Ratnaparkhi, 1998).

Second, we observe that the lexical and sublexical features we counted are not sufficient to identify all the relevant linguistic classes: statistical tests fail to differentiate between unaccusatives and object-drop verbs. In order to distinguish between these two classes of verbs one needs to look at some of the surrounding context. This result is expected. Performance measures of statistical parsers show that statistics based on one word give poor results, but that statistics on bigrams have much better performance (Charniak, 1997).

3 General Discussion

3.1 Relationship between Different Kinds of Methods

Our results cast some light on an important methodological question: can frequencies in annotated corpora be considered a good approximation of speakers' preferences? Recent results in the literature have argued that they cannot, showing large discrepancies between data collection methods (Merlo, 1994), comprehension and production (Gibson et al., 1996), and online preferences and corpora counts (Brysbaert et al., 1998). Several explanations have been proposed, mostly dismissive of some particular method to collect data; for example: frequency-based preferences are not used by humans; the wrong frequencies had been counted; experimental results are not representative of natural linguistic behaviour; or corpora are not representative of natural linguistic behaviour. The findings in this study show a way of reconciling results obtained by different data collection methods: if we count at the level of lexical and sublexical features, we find that differences in native speakers' preferences do correspond to significant differences in distributions. Similar conclusions are being reached in (Roland and Jurafsky, 1998), who compare different corpora.

3.2 Classification Properties of Lexical Features and Consequences

Looking at the frequencies of the lexical features in Table 2, we can observe that PRT, PASS and TRANS have counts that can be used to directly predict the difficulty of the RR construction. This observation can be used beneficially in a task different from parsing, for instance in a generation system. Some current methods have a generate and filter approach (Knight and Hatzivassiloglou, 1995): all constructions are generated and then filtered based on a statistical model. If the trigram model has a good fit with text, our experiments indicates that it would eliminate many RRs for unaccusatives that would be considered acceptable by speakers. If instead the filtering is based on, for example, the frequency of the past participle use, the system would correctly allow unaccusative RRs, but filter out unergative RRs.

Moreover, we notice that all the lexical features reproduce the well-known relation between markedness within a language and typology of languages: what is an existing but infrequent construction in a few languages is absent in many languages. In this instance, the transitive use of manner of motion verbs — *The rider raced the horse* — is a marked construction in English, in the sense that while it is grammatical, this use is only restricted to a subset of manner of motion verbs. This construction which is marked in English is ungrammatical in Romance; languages such as Italian or French do not have a grammatical direct translation for the sentence above. This is called in the social sciences a *zero-rare* distribution, where a feature that is generally already rare is however *never* present in one subclass of the cases.

Interestingly, the lexical feature ADJ presents a distribution that reflects this cross-linguistic fact internally to English: unergative verbs never occur prenominally, even those that can occur transitively, passively and in reduced relative clauses.

This is a particularly useful distributional cue for verb classification. On observing the complete absence of prenominal adjectives derived from transitive verbs one can classify the verb as unergative. Or, the cue provided can be used in a translation task: one of the typical argument structure divergences between English and Romance languages can be inferred by looking at distributional data. Thus, by observing the absence of prenominal adjectives in English the translation system can avoid proposing the RR alternative in the target language, where it would be ungrammatical.

3.3 Language Engineering

Finally, this kind of in-depth corpus analysis gives us indications on what kind of syntactic annotation is needed in order to be able to use a corpus to perform tasks at the sentence level, and also, possibly, how to bootstrap a syntactic annotation process in a way that does not require much in-depth semantic knowledge about words.

Had we wanted to perform the study reported in this paper by simple counting of occurrences in an appropriately annotated text — thus eliminating the need for the tedious and time-consuming filtering of the automatic extraction which was necessary in the present study — we would have needed a text annotated with categories derived from knowledge about individual lexical items and a small portion of the tree surrounding them. First, all our counts assumed knowledge of the verb classification in unergative, unaccusative and object-drop, which requires annotation of the thematic roles of the verb. Furthermore, for the counts of the several variables described we needed the verb items and the preceding auxiliary (active-passive and MV/past participle), the following noun phrase and knowledge about whether the noun phrase was the direct object of the verb or not (transitive-intransitive), the preceding noun phrase and knowledge about whether the noun phrase was the subject of the verb or rather an adjunct head (MV/reduced relative), and the preceding determiner (adjective-non-adjective). This is evidence in favour of annotation using a lexicalised formalism, whose main units are argument-structure dependencies between words, whether encoded structurally, as in LTAG (Schabes and Joshi, 1991), or as grammatical relations, as in dependency grammar (Hudson, 1990; Mel'cuk, 1988). From the point of view of parsing, these counts require only one chunk of text each.

As an example, consider a grammatical formalisms, such as LTAG (Schabes and Joshi, 1991), which is both lexicalised and has been used to chunk text without performing a full parse. An LTAG lexicon is a forest of lexicalised elementary trees. For verbs, the tree structure corresponds to their argument structure. Thus, each of the lexical items and portion of tree mentioned above correspond to a different elementary tree, including the unergative and unaccusative distinction, encoded by different labels referring to thematic roles. Current

LTAG part-of-speech taggers, called supertaggers (Joshi and Srinivas, 1994; Srinivas, 1997) assign a set of elementary trees to each word, in effect chunking the text. The counts performed in the study reported here would have required simply counting the occurrences of the labels assigned to the words in the text by such a supertagger. Refinements in this direction of the annotation of the grammar used by the XTAG system (Doran et al., 1994) are actually under way.

We also can see, from the raw frequencies obtained, that when collecting counts about syntactic phenomena, corpora must be in the order of hundreds of millions of words for the statistics to be reliable.

4 Conclusions

Our main result in this paper is that statistics over *lexical* features best correspond to independently established human intuitive judgments. We have argued that, methodologically, this result casts light on the relationship between different data collection methods, and shows that some apparently contradictory results can be reconciled by defining probability spaces at the lexical and sublexical level. From the point of view of language engineering, we have argued that this result provides an indication of what units might reflect preferences that port across tasks, and what type of syntactic annotation of corpora is going to be most useful.

5 Acknowledgments

This research was partly sponsored by the Swiss National Science Foundation, under fellowship 8210-46569 to P. Merlo, and by the US National Science Foundation, under grant #9702331 to S. Stevenson. We thank Aravind Joshi, Martha Palmer and Adwait Ratnaparkhi for useful comments.

References

Thomas G. Bever. 1970. The cognitive basis for linguistic structure. In J. R. Hayes, editor, *Cognition and the Development of Language*. John Wiley, New York.

M. Brysbaert, D.C. Mitchell, and Stefan Grondelaers. 1998. Cross-linguistic differences in modifier attachment biases: Evidence against Gricean and Tuning accounts. manuscript.

Eugene Charniak. 1997. Statistical parsing with a context-free grammar and word statistics. In *Proc. of the 14th National Conference on AI.*

Michael John Collins. 1997. Three generative, lexicalised models for statistical parsing. In *Proc. of the 35th Annual Meeting of the ACL*, pages 16–23.

Christy Doran, Dania Egedi, Beth Ann Hockey, B. Srinivas, and Martin Zaidel. 1994. XTAG system - a wide coverage grammar for English. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)*, pages 922–928, Kyoto, Japan.

E. Gibson, C. Schütze, and A. Salomon. 1996. The relationship between the frequency and the processing complexity of linguistic structure. *J. of Psych. Research*, 25(1):59–92.

Ken Hale and Jay Keyser. 1993. On argument structure and the lexical representation of syntactic relations. In K. Hale and J. Keyser, editors, *The View from Building 20*, pages 53–110. MIT Press.

Richard Hudson. 1990. *English Word Grammar.* Basil Blackwell.

A. Joshi and B. Srinivas. 1994. Disambiguation of super parts of speech (supertags): Almost parsing. In *Proc. of Coling 94*, Kyoto, Japan.

Kevin Knight and Vasileios Hatzivassiloglou. 1995. Two-level, many-paths generation. In *Proc. of the 33th Annual Meeting of the ACL*, pages 252–260, Cambridge,MA.

Beth Levin and Malka Rappaport Hovav. 1995. *Unaccusativity.* MIT Press, Cambridge, MA.

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

Maryellen C. MacDonald, Neal J. Pearlmutter, and Mark Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review.*

Maryellen MacDonald. 1994. Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9(2):157–201.

I. Mel'cuk. 1988. *Dependency Syntax: Theory and practice.* SUNY Press, Albany.

Paola Merlo. 1994. A corpus-based analysis of verb continuation frequencies for syntactic processing. *Journal of Psycholinguistics Research*, 23(6):435–457.

Adwait Ratnaparkhi. 1996. A maximum entropy part-of-speech tagger. In *Proceedings of the Empirical Methods in Natural Language Processing Conference*, Philadelphia, PA. University of Pennsylvania.

Adwait Ratnaparkhi. 1997. A linear observed time statistical parser based on maximum entropy models. In *2nd Conf. on Empirical Methods in NLP*, pages 1–10, Providence, RI.

Adwait Ratnaparkhi. 1998. Statistical models for unsupervised prepositional phrase attachment. In *Proc. of the 36th Annual Meeting of the ACL*, Montreal, CA.

Doug Roland and Dan Jurafsky. 1998. How verb subcategorization frequencies are affected by corpus choice. In *Proc. of the 36th Annual Meeting of the ACL*, Montreal, CA.

Yves Schabes and Aravind Joshi. 1991. Parsing with lexicalized tree adjoining grammars. In Masaru Tomita, editor, *Current Issues in Parsing Technology*. Kluwer Academic Publishers.

B. Srinivas. 1997. *Complexity of Lexical Descriptions and its Relevance to Partial Parsing*. Ph.D. thesis, University of Pennsylvania.

Suzanne Stevenson and Paola Merlo. 1997. Lexical structure and processing complexity. *Language and Cognitive Processes*.

John Trueswell, Michael Tanenhaus, and Susan Garnsey. 1994. Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33:285–318.

John Trueswell. 1996. The role of lexical frequency in syntactic ambiguity resolution. *J. of Memory and Language*, 35:566–585.