Language Acquisition and Ambiguity Resolution: The Role of Frequency Distributions

Paola Merlo
LATL-University of Geneva
Department of Linguistics
2 rue de Candolle
1211 Genève 4
Switzerland
merlo@lettres.unige.ch

Suzanne Stevenson
Department of Computer Science
and Center for Cognitive Science (RuCCS)
Rutgers University
110 Frelinghuysen Road
New Brunswick, NJ 08854-8019
suzanne@cs.rutgers.edu

Abstract

This paper proposes that the set of frequencies that the human language processor keeps track of are those that are useful to it in learning. In a computational experimental setting, we investigate four linguistically motivated features which distinguish subclasses of intransitive verbs, and suggest that those features that are the most useful to automatically classify verbs into lexical semantic classes are related to mechanisms used in adult processing to resolve structural ambiguity.

Introduction

Models of human language comprehension have traditionally focused on discrete linguistic properties—structural or interpretive factors—as the guiding influence in determining the preferred interpretation of an ambiguity (e.g., [8, 18]). Theories of human sentence processing that are founded on such linguistic distinctions have generally assumed that their use is an inherent (and universal) property of the language processor. Even if certain distinctions must be learned (such as those that involve parametric variation among languages), the learning process itself has been irrelevant to the later use of those features in resolving ambiguity.

Recently, the experience-based paradigm has shifted the emphasis to the role of frequencies—continuously-valued information that weights the contribution of individual linguistic and contextual features within the ambiguity resolution process [9, 15, 17, 25]. Clearly, frequencies are acquired through on-going exposure to linguistic input, and so these approaches promise to integrate learning and processing more closely. While some work has shown that properties of a learning mechanism may underlie difficulty in processing (e.g., in the case of embedded structures, as in [4]), to our knowledge none has explored the connection between properties of the learning process and the features that play a role in guiding ambiguity resolution.

Here we extend the experience-based point of view by proposing that the frequency distributions that play a guiding role in ambiguity resolution are those that contribute to language acquisition. We assume that the language learner must learn certain fundamental linguistic distinctions. (Not all can be innate, since at least some are lexically specific.) We further assume that distributional data (the frequencies of alternative features and/or constructions) play a role in the learning process. We view the adult language processor as a mature version of the child language processor, with access to the type of knowledge that guides acquisition of language, including the distributional information. Since different resolutions of an ambiguity may be distinguished by fundamental linguistic factors that must be learned, we expect that the frequency distributions of those factors will contribute to the process of selecting the preferred interpretation.

Our general hypothesis then is that the distributional knowledge used by the language processor in guiding interpretation is restricted to specific features that are required to learn necessary linguistic distinctions. Alternatively stated, the frequencies used in language comprehension are a subset (not necessarily a proper subset) of the frequencies used in language acquisition.

Verb Classes and Ambiguity Resolution

We investigate our general hypothesis by exploring the specific instance of the relation between the learning of lexical semantic verb classifications from distributional data, and the use of those frequency distributions in ambiguity resolution. We focus on the main verb/reduced relative (MV/RR) ambiguity [1]:

(1) The horse **raced** past the barn fell.
(2) The boy **washed** in the tub was angry.

In (1) and (2), the boldfaced verb can be interpreted as either a past tense main verb, or as a past participle within a reduced relative clause (e.g., *the horse* [that was] *raced past the barn*). In each case, the reduced relative interpretation is required for a coherent analysis of the complete sentence. The main verb interpretation of *raced* is so strongly preferred that people have great difficulty understanding sentence (1). However, the ease of sentence (2) shows that the reduced relative interpretation is not difficult for all verbs.

The differences in ease of interpreting the resolutions of this ambiguity have been shown to be sensitive both to frequency differentials [16, 26] and to verb class distinctions [24, 7]. Within the context of our general hypothesis above, we relate these two factors by proposing that the frequency measures that influence the ambiguity resolution process in this construction are intimately related to the defining properties of three classes of verbs that present this ambiguity. Building on the idea of syntactic bootstrapping [10], and on statistical approaches to extracting verb information from frequency distributions [12, 2, 23], we follow a computational experimental methodology in which we investigate as indicated each of the following specific hypotheses:

H1: Linguistically motivated features for distinguishing the verb classes are apparent within linguistic experience.

> We analyze the three classes to determine potentially relevant distinctive features, and count those features (or approximations to them) in a very large corpus.

H2: The distributional patterns of (some of) those features contribute to learning the classifications of the verbs.

> We apply machine learning techniques to determine whether the features support the learning of the classifications. We analyze the contribution of different features to the classification process.

H3: Features informative in learning the verb classes also influence the resolution of the MV/RR ambiguity, and features not helpful in learning do not affect processing of the ambiguity.

> We examine whether the features that are found to be informative in classification play a role in resolution of the ambiguity. Conversely, we discuss whether features that are not found to contribute to learning similarly play no role in processing.[1]

To preview, we find that, related to (H1), linguistically motivated features that distinguish the verb classes can be extracted from a corpus with a moderate amount of annotation. We assume that these features are available to the learner once it can make certain fundamental syntactic distinctions (e.g., part of speech, constituency). In relation to (H2), a subset of these features is sufficient to halve the error rate compared to chance in automatic verb classification, suggesting that distributional data does in fact contribute to our knowledge of the classification of verbs. Furthermore, we find that features that are distributionally predictable, because they are highly correlated to other features, contribute little to classification performance. We conclude that the usefulness of distributional features to the learner is determined by their linguistic informativeness. Finally, we find that the results in (H3) provide important preliminary evidence concerning our general hypothesis that the set of frequency differentials that are available to the learning algorithm and the processing algorithm are the same. Features which contribute most to learning the verb classes have been experimentally demonstrated to influence processing of the MV/RR ambiguity, and one which is not informative for learning has no evidence for its role in MV/RR ambiguity resolution.

[1]Note that the latter is not directly testable, since it may be that current experiments have simply not shown the effect of the variable even though it does influence processing. Thus any conclusions here are suggestive only.

Determining the Features

In this section, we present motivation for the features that we investigate in terms of their role in learning the verb classes. We first present the linguistically derived features, then turn to an analysis of the MV/RR ambiguity to extend the set of potentially relevant features.

Features of the Verb Classes

The MV/RR ambiguity involves a choice between a main verb form that is intransitive, and a reduced relative form that is transitive (because the reduced relative is a passive use of the verb). The three verbs classes under study—unergative, unaccusative, and object-drop—were thus chosen because they exhaustively partition the optionally intransitive verbs in English. The three verb classes differ in the properties of their intransitive/transitive alternations, which are exemplified below.

Unergative:

(3a) The horse raced past the barn.

(3b) The jockey raced the horse past the barn.

Unaccusative:

(4a) The butter melted in the pan.

(4b) The cook melted the butter in the pan.

Object-drop:

(5a) The boy washed the hall.

(5b) The boy washed.

The sentences in (3) use an unergative verb, *raced*. Unergatives are intransitive action verbs whose transitive form is the causative counterpart of the intransitive form. Thus, the subject of the intransitive (3a) becomes the object of the transitive (3b) [3, 11, 14]. The sentences in (4) use an unaccusative verb, *melted*. Unaccusatives are intransitive change of state verbs (4a); like unergatives, the transitive counterpart for these verbs is also causative (4b). The sentences in (5) use an object-drop verb, *washed*; these verbs have a non-causative transitive/intransitive alternation, in which the object is simply optional.

Both unergatives and unaccusatives have a causative transitive form, but differ in the semantic roles that they assign to the participants in the event described. In an intransitive unergative, the subject is an Agent (the doer of the event), and in an intransitive unaccusative, the subject is a Theme (something affected by the event).

The role assignments to the corresponding semantic arguments of the transitive forms—i.e., the direct objects—are the same, with the addition of a Causal Agent (the causer of the event) as subject in both cases. This leads to an unusual situation for a transitive unergative, because it assigns two agentive roles—the subject is the agent of causation, and the object is the agent of the action expressed by the verb [24]. Object-drop verbs have a simpler participant/role mapping than either unergatives or unaccusatives, assigning Agent to the subject and Theme to the optional object.

We expect the differing semantic role assignments of the verb classes to be reflected in their syntactic behavior [13, 6], and consequently in the distributional data we collect from a corpus. The three classes can be characterized by their occurrence in two alternations: the intransitive/transitive alternation and the causative alternation. Unergatives are distinguished from the other classes in being rare in the transitive form (due to their "double agentive" nature); we expect this to be reflected in a greater degree of intransitive use in a corpus. Both unergatives and unaccusatives are distinguished from object-drop in being causative in their transitive form, and similarly we expect this to be reflected in amount of detectable causative use. Furthermore, since the causative is a transitive use, and the transitive use of unergatives is expected to be rare, causativity should primarily distinguish unaccusatives from object-drops. In conclusion, we expect the defining features of the verb classes—the intransitive/transitive and causative alternations—to lead to distributional differences in the observed usages of the verbs in these alternations.

Features of the MV/RR Alternatives

We now examine the features that distinguish the two resolutions of the MV/RR ambiguity:

Main Verb: The horse raced past the barn quickly.

Reduced Relative: The horse raced past the barn fell.

In the main verb resolution, the ambiguous verb *raced* is used in its intransitive form, while in the reduced relative, it is used in its transitive, causative form. These features correspond directly to the defining alternations of the three verb classes under study (intransitive/transitive, causative). Additionally, we see that other, related features serve to distinguish the two resolutions of the ambiguity. The main verb form is active and a main verb part-of-speech (tagged as VBD); by contrast, the reduced relative form is passive and a past participle (tagged as VBN). Note that these properties are redundant with the intransitive/transitive distinction, as passive implies transitive use, and necessarily entails the use of a past participle.

We add the VBD/VBN and active/passive distinctions to our features for classification (the transitive and causative alternations) motivated by two factors. First, recent work in machine learning [21, 22] has argued that using overlapping features can be beneficial for learning. Second, one of the features—the VBD/VBN distinction—has already been shown to influence resolution of the MV/RR ambiguity [26].

In the next section, we describe how we compile the corpus counts for each of the four properties, in order to approximate the distributional information of these alternations available to a human learner.

Frequency Distributions of the Features

We assume that currently available large corpora are a reasonable approximation to the linguistic input that the learner is exposed to [19]. Using a combined corpus of 65-million words, we measured the relative frequency distributions of the linguistic features (VBD/VBN, active/passive, intransitive/transitive, causative) over a sample of verbs from the three lexical semantic classes.

We chose a total of 60 verbs, 20 verbs from each class, based on the classification of verbs in [13] (see Appendix A). Each verb presents the same form in the simple past and in the past participle, as in the MV/RR ambiguity, and most of the verbs can occur in the transitive and in the passive. Most counts were performed on the Linguistic Data Consortium release of the Brown Corpus and of years 1987, 1988, 1989 of the Wall Street Journal, a combined corpus in excess of 65 million words labeled with part-of-speech (POS) tags.[2] Due to the need for additional annotation, the causative feature was counted only for the 1988 year of the WSJ, a parsed corpus of 29 million words also available from the LDC (parsed with the parser from [5]).

The counts for VBD/VBN were automatically extracted for all verbs based on the part-of-speech label according to the tagged corpus. For intransitive/transitive, we searched for the closest nominal group following the verb which was considered to be the object. For active/passive uses, we looked for the preceeding auxiliary: *have* indicates an active use and *be* indicates a passive use. The causative feature was approximated by extracting heads of subjects and objects for each verb occurrence, and calculating a token-based percentage of overlap between the two sets of nouns. This captures the property of the causative construction that the subject of the intransitive can occur as the object of the transitive. All counts were normalized, yielding a total of four relative frequency features: VBD (%VBD tag), ACT (%active use), INTR (%intransitive use), CAUS (%causative use). All raw and normalized corpus data are available from

[2] Because we need a very large corpus due to the constraints imposed by the statistical methods used, we are restricted at the present time primarily to newspaper text of this type. Clearly, verificational work across different kinds of corpora, including child-directed speech, would be helpful in elaborating our proposal.

Features	Accuracy
1. VBD ACT INTR CAUS	52%
2. VBD INTR CAUS	66%
3. ACT INTR CAUS	47%
4. VBD ACT CAUS	54%
5. VBD ACT INTR	45%

Table 1: Accuracy of the Verb Clustering Task.

	Decision Trees		Rule Sets	
Features	A%	SE%	A%	SE%
1. VBD ACT INTR CAUS	64.2	1.7	64.9	1.6
2. VBD INTR CAUS	60.9	1.2	62.3	1.2
3. ACT INTR CAUS	59.8	1.2	58.9	0.9
4. VBD ACT CAUS	55.4	1.5	55.7	1.4
5. VBD ACT INTR	54.4	1.4	56.7	1.5

Table 2: Accuracy (A%) and Standard Error (SE%) in the Verb Classification Task.

the authors.

Frequencies in Learning and Processing

The frequency distributions of the verb alternation features yield a vector for each verb that represents the relative frequency values for the verb on each dimension; the set of 60 vectors constitute the data for our machine learning experiments.

Vector template: [verb-name, VBD, ACT, INTR, CAUS]

Example: [opened, .793, .910, .308, .158]

It must be determined experimentally which of the distributions actually contribute to learning the verb classifications. First we describe computational experiments in unsupervised learning, using hierarchical clustering, then we turn to supervised learning, using decision tree induction. We conclude with a discussion of the relation of informative features in learning and processing.

Unsupervised Learning

We used the hierarchical clustering algorithm available in SPlus5.0, imposing a cut point that produced three clusters, to correspond to the three verb classes. Table 1 shows the accuracy achieved using the four features that discriminate the resolutions of the MV/RR ambiguity (row 1), and all three-feature subsets of those four features (rows 2–5). Note that chance performance in this task (a three-way classification) is 33% correct.

The highest accuracy in clustering, of 66%—or half the error rate compared to chance—is obtained only by the triple of features in row 2 in the table: VBD, INTR, and CAUS. All other subsets of features yield a much lower accuracy, of 45-54%. We can conclude that some of the features contribute useful information to guide clustering, but the inclusion of ACT actually degrades performance. Clearly, having fewer but more relevant features is important to accuracy in verb classification. We will return to the issue of which features contribute most to learning in our discussion of supervised learning below.

A problem with analyzing the clustering performance is that it is not always clear what counts as a misclassification. We cannot actually know what the identity of the verb class is for each cluster. In the above results, we imposed a classification based on the class of the majority of verbs in a cluster, but a problem arises when there is no clear majority. The determination of the number of cut points introduces a further degree of uncertainty in interpreting the results.

To evaluate better the effects of the features in learning, we therefore turned to a supervised learning method, in which the algorithm trains on cases with known classification. We do not believe that supervised learning is realistic for the human learner in this case. Rather, we perform these computational experiments in order to measure, on clear classification results, the contribution in principle of the features to learning.

Supervised Learning

For our supervised learning experiments, we used the publicly available version of the C5.0 machine learning algorithm (http://www.rulequest.com/), a newer version of C4.5 [20]. Given a training set with known classifications, this system applies inductive learning to create a procedure that can be applied to label new cases whose classification is unknown. The output of the system is presented in the form of either decision trees or rule sets, alternative data structures for encoding the use of the features in classification. For all reported experiments, we ran a 10-fold cross-validation repeated ten times, and the numbers reported are averages over all the runs.[3]

Table 2 show the results of our experiments on (different combinations of) the four features we counted in the corpus. Recall that chance performance in this task (a three-way classification) is 33% correct. We attain the best performance, of 64–65% (again, almost halving the chance error rate), by using all four features that discriminate the resolutions of the MV/RR ambiguity: VBD, ACT, INTR, and CAUS (row 1 in the table).

Comparing the accuracy of each of the three-feature subsets to the four-feature subset is very informative.

[3] A 10-fold cross-validation means that the system randomly divides the data into ten parts, and runs ten times on a different 90%-training-data/10%-test-data split, yielding an average accuracy and standard error. This procedure is then repeated for 10 different random divisions of the data, and accuracy and standard error are again averaged across the ten runs.

When either the INTR or CAUS feature is removed (rows 4 and 5 respectively), performance degrades considerably, with a decrease in accuracy of 8–10% from the maximum achieved with the four features. However, when the VBD feature is removed (row 3), there is a smaller decrease in accuracy, of 4–6%. When the ACT feature is removed (row 2), there is an even smaller decrease, of 2–4%. In fact, taking the standard error into account, the accuracy using all features (row 1) is not distinct from using only the three features VBD, INTR, and CAUS (row 2)—the same subset of features that yielded the best performance in clustering. We conclude then that INTR and CAUS contribute most to the accuracy of the classification, while ACT seems to contribute little. This shows that not all the linguistically relevant features are equally useful in learning.

We think that this pattern of results is related to the combination of the feature distributions: some distributions are highly correlated, while others are not. According to our calculations, CAUS is not significantly correlated with any other feature. Of the features that are significantly correlated, ACT and VBD are the most highly correlated (R=.67), ACT and INTR the next most highly correlated (R=.44), and VBD and INTR the least correlated (R=.36). Thus, CAUS is uncorrelated with the other features, and VBD and INTR are the least correlated pair of the remaining three.

We expect combinations of features that are less correlated to yield better classification accuracy. If we compare the accuracy of the 3-feature combinations in Table 2 (rows 2–5), this hypothesis is confirmed. The three combinations that contain the feature CAUS (rows 2, 3 and 4)—the uncorrelated feature—have better performance than the combination that does not (row 5), as expected. Now consider the subsets of three features that include CAUS with a pair of the other correlated features. The combination containing VBD and INTR (row 2)—the least correlated pair of features—has the best accuracy, the combination containing ACT and INTR (row 3)—the next least correlated pair—has the next highest accuracy, while the combination containing the most highly correlated ACT and VBD (row 4) has the worst accuracy. Thus, redundancy of features appears to play a clear role in determining their usefulness in learning.[4]

Comparison of Learning and Processing

Recall our motivating hypothesis that only those features that are useful in acquisition are available to influence interpretation. Given our learning results, then, we expect to see that INTR, CAUS, and VBD play a role in ambiguity resolution, while ACT does not. Currently available experimental data yield preliminary support for this conclusion. Relative frequency of VBD/VBN use [26] and intransitive/transitive use [16] have been shown to influence resolution of the MV/RR ambiguity.

Demonstrating a processing effect of the frequency of causativity is more problematic. For unergatives, we note that there is an ineliminable confound between verb class and distribution of the CAUS feature. Unergatives never occur as causatives in our very large corpus, thus indicating that the causative use of unergatives is very rare. We interpret this fact as indicating that the causative use of an unergative is difficult, even though linguistically possible, because it entails a double agent structure, as discussed above [24]. Whether caused by knowledge of verb class directly or by causativity, effects of extreme difficulty have been found for unergatives in the MV/RR ambiguity.

On the other hand, unaccusatives occur more often in a causative construction, with a gradient of use. Recently, [7] have manipulated the degree of proto-agent properties, which include causativity, in the MV/RR ambiguity. They collected subject ratings, and showed that the ratings of these properties can influence the ease or difficulty of the reduced relative interpretation. While it remains to be shown that proto-agent ratings reflect the causative frequency distribution in a corpus, this data is highly suggestive.

By contrast to the INTR, CAUS, and VBD features, there is no experimental data indicating that ACT plays a role in MV/RR interpretation (or, in fact, in the resolution of any other ambiguity). Although these results are preliminary, they lend initial support within this constrained domain for our general hypothesis regarding the relation between learning and processing.

Conclusions

In this paper, we propose that the human language processor keeps track of the set of frequencies that are useful to it in learning. To explore this general hypothesis within a specific domain, we investigate how a set of linguistic features that distinguish intransitive verbs perform in a classification task. Linguistic analysis predicts certain features are most relevant (INTR and CAUS), but learning is improved by also using a partially redundant feature (VBD). Experimental evidence in sentence processing shows that some of these features, such as INTR and VBD, are used in resolving ambiguity. Others, such as CAUS, have been indirectly shown to have an influence on ambiguity resolution. On the other hand, a highly redundant feature, ACT, is not informative for learning. Interestingly, there is no experimental data indicating that ACT plays a role in syntactic disambiguation. Our proposal thus receives preliminary support, and also makes a

[4] We suspect that another factor comes into play, namely how noisy the feature is. The similarity in performance using INTR or CAUS in combination with VBD and ACT (rows 4 and 5) might also be due in part to the fact that the counts for CAUS are a more noisy approximation of the actual feature distribution than the counts for INTR. We reserve defining a precise model of noise, and its interaction with the other factors, for future research.

simple prediction which is easily testable, that ACT is not useful in disambiguating verb alternation ambiguities.

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 grants #9702331 and #9818322 to S. Stevenson. We thank Martha Palmer for getting us started on this work, and Michael Collins for giving us acces to the output of his parser.

Appendix A

The unergatives are manner of motion verbs: *jumped, rushed, marched, leaped, floated, raced, hurried, wandered, vaulted, paraded, galloped, glided, hiked, hopped, jogged, scooted, scurried, skipped, tiptoed, trotted.*

The unaccusatives are verbs of change of state: *opened, exploded, flooded, dissolved, cracked, hardened, boiled, melted, fractured, solidified, collapsed, cooled, folded, widened, changed, cleared, divided, simmered, stabilized.*

The object-drop verbs are unspecified object alternation verbs: *played, painted, kicked, carved, reaped, washed, danced, yelled, typed, knitted, borrowed, inherited, organized, rented, sketched, cleaned, packed, studied, swallowed, called.*

References

[1] Thomas G. Bever. The cognitive basis for linguistic structure. In J. R. Hayes, editor, *Cognition and the Development of Language*, pages 278–352. John Wiley, NY, 1970.

[2] Michael Brent. From grammar to lexicon: Unsupervised learning of lexical syntax. *Computational Linguistics*, 19(2):243–262, 1993.

[3] Anne-Marie Brousseau and Elizabeth Ritter. A non-unified analysis of agentive verbs. In Dawn Bates, editor, *West Coast Conference on Formal Linguistics*, number 20, pages 53–64, Stanford, CA, 1991. Center for the Study of Language and Information.

[4] Morten H. Christiansen and Maryellen C. MacDonald. Individual differences in sentence comprehension: The importance of experience. Talk at the 11th Annual CUNY Conference on Human Sentence Processing, 1998.

[5] Michael Collins. Three generative, lexicalized models for statistical parsing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pages 16–23, 1997.

[6] Hoa Trang Dang, Karin Kipper, Martha Palmer, and Joseph Rosenzweig. Investigating regular sense extensions based on interesective Levin classes. In *Proc. of the 36th Annual Meeting of the ACL (COLING-ACL '98)*, pages 293–299, 1998.

[7] Hana Filip, Michael K. Tanenhaus, Gregory N. Carlson, Paul D. Allopenna, and Joshua Blatt. Reduced relatives judged hard require constraint-based analyses. In Paola Merlo and Suzanne Stevenson, editors, *Sentence Processing and the Lexicon: Formal, Computational and Experimental Perspectives*. John Benjamins, Holland, 1999.

[8] Lyn Frazier. *On Comprehending Sentences: Syntactic Parsing Strategies*. PhD thesis, University of Connecticut, 1978. Available through the Indiana University Linguistics Club, Bloomington, IN.

[9] Susan M. Garnsey, Neal J. Pearlmutter, Elizabeth Myers, and Melanie A. Lotocky. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37:58–93, 1997.

[10] Lila Gleitman. The structural sources of verb meaning. *Language Acquisition*, 1:3–56, 1990.

[11] Ken Hale and Jay Keyser. On argument structure and the lexical representation of syntactic relations. In K. Hale and J. Keyser, editors, *The View from Building 20*, pages 53–109. MIT Press, Cambridge, MA, 1993.

[12] Judith L. Klavans and Martin Chodorow. Degrees of stativity: The lexical representation of verb aspect. In *Proceedings of the Fourteenth International Conference on Computational Linguistics*, 1992.

[13] Beth Levin. *English Verb Classes and Alternations*. University of Chicago Press, Chicago, IL, 1993.

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

[15] Maryellen MacDonald, Neal Pearlmutter, and Mark Seidenberg. Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4):676–703, 1994.

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

[17] D. C. Mitchell, F. Cuetos, M. M. B. Corley, and M. Brysbaert. Exposure-based models of human parsing: Evidence for the use of coarse-grained (non-lexical) statistical records. *J. of Psycholinguistic Res.*, 24:469–488, 1995.

[18] Bradley Pritchett. *Grammatical Competence and Parsing Performance*. University of Chicago Press, 1992.

[19] Geoffrey K. Pullum. Learnability, hyperlearning, and the poverty of the stimulus. In Jan Johnson, Matthew L. Juge, and Jeri L. Moxley, editors, *22nd Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on the Role of Learnability in Grammatical Theory*, pages 498–513, Berkeley, California, 1996. Berkeley Linguistics Society.

[20] J. Ross Quinlan. *C4.5 : Programs for Machine Learning*. Series in Machine Learning. Morgan Kaufmann, San Mateo, CA, 1992.

[21] Adwait Ratnaparkhi. A linear observed time statistical parser based on maximum entropy models. In *Proceedings of the 2nd Conference on Empirical Methods in NLP*, pages 1–10, 1997. Providence, RI.

[22] Adwait Ratnaparkhi. Statistical models for unsupervised prepositional phrase attachment. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics*, pages 1079–1085, 1998.

[23] Philip Resnik. Selectional constraints: an information-theoretic model and its computational realization. *Cognition*, 61(1–2):127–160, 1996.

[24] Suzanne Stevenson and Paola Merlo. Lexical structure and parsing complexity. *Language and Cognitive Processes*, 12(2/3):349–399, 1997.

[25] John Trueswell and Michael J. Tanenhaus. Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In Charles Clifton, Lyn Frazier, and Keith Rayner, editors, *Perspectives on Sentence Processing*, pages 155–179. Lawrence Erlbaum, NJ, 1994.

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