SEMANTIC CLASSES AND SYNTACTIC AMBIGUITY

*Philip Resnik**

Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104
resnik@linc.cis.upenn.edu

ABSTRACT

In this paper we propose to define selectional preference and semantic similarity as information-theoretic relationships involving conceptual classes, and we demonstrate the applicability of these definitions to the resolution of syntactic ambiguity. The space of classes is defined using WordNet [8], and conceptual relationships are determined by means of statistical analysis using parsed text in the Penn Treebank.

1. INTRODUCTION

The problem of syntactic ambiguity is a pervasive one. As Church and Patil [2] point out, the class of "every way ambiguous" constructions — those for which the number of analyses is the number of binary trees over the terminal elements — includes such frequent constructions as prepositional phrases, coordination, and nominal compounds. They suggest that until it has more useful constraints for resolving ambiguities, a parser can do little better than to efficiently record all the possible attachments and move on.

In general, it may be that such constraints can only be supplied by analysis of the context, domain-dependent knowledge, or other complex inferential processes. However, we will suggest that in many cases, syntactic ambiguity can be resolved with the help of an extremely limited form of semantic knowledge, closely tied to the lexical items in the sentence.

We focus on two relationships: selectional preference and semantic similarity. From one perspective, the proposals here can be viewed as an attempt to provide new formalizations for familiar but seldom carefully defined linguistic notions; elsewhere we demonstrate the utility of this approach in linguistic explanation [11]. From another perspective, the work reported here can be viewed as an attempt to generalize statistical natural language techniques based on lexical associations, using knowledge-based rather than distributionally derived word classes.

2. CLASS-BASED STATISTICS

A number of researchers have explored using lexical co-occurrences in text corpora to induce word classes [1, 5, 9, 12], with results that are generally evaluated by inspecting the semantic cohesiveness of the distributional classes that result. In this work, we are investigating the alternative of using Word-Net, an explicitly semantic, broad coverage lexical database, to define the space of semantic classes. Although WordNet is subject to the attendant disadvantages of any hand-constructed knowledge base, we have found that it provides an acceptable foundation upon which to build corpus-based techniques [10]. This affords us a clear distinction between domain-independent and corpus-specific sources of information, and a well-understood taxonomic representation for the domain-independent knowledge.

Although WordNet includes data for several parts of speech, and encodes numerous semantic relationships (meronymy, antonymy, verb entailment, etc.), in this work we use only the noun taxonomy — specifically, the mapping from words to word classes, and the traditional IS-A relationship between classes. For example, the word *newspaper* belongs to the classes ⟨newsprint⟩ and ⟨paper⟩, among others, and these are immediate subclasses of ⟨material⟩ and ⟨publisher⟩, respectively.[1]

Class frequencies are estimated on the basis of lexical frequencies in text corpora. The frequency of a class c is estimated using the lexical frequencies of its members, as follows:

$$\text{freq}(c) = \sum_{\{n \mid n \text{ is subsumed by } c\}} \text{freq}(n) \qquad (1)$$

The class probabilities used in the section that follows can then be estimated by simply normalizing (MLE) or by other methods such as Good-Turing [3].[2]

---

[1]For expository convenience we identify WordNet noun classes using a single descriptive word in angle brackets. However, the internal representation assigns each class a unique identifier.

[2]We use Good-Turing. Note, however, that WordNet classes are not necessarily disjoint; space limitations preclude further discussion of this complication here.

3. CONCEPTUAL RELATIONSHIPS

3.1. Selectional Preference

The term "selectional preference" has been used by linguists to characterize the source of anomaly in sentences such as (1b), and more generally to describe a class of restrictions on co-occurrence that is orthogonal to syntactic constraints.

(1)   a. John admires sincerity.
         b. Sincerity admires John.

(2)   a. Mary drank some wine.
         b. Mary drank some gasoline.
         c. Mary drank some pencils.
         d. Mary drank some sadness.

Although selectional preference is traditionally formalized in terms of feature agreement using notations like [+Animate], such formalizations often fail to specify the set of allowable features, or to capture the gradedness of qualitative differences such as those in (2).

As an alternative, we have proposed the following formalization of selectional preference [11]:

Definition. The *selectional preference* of w for C is the relative entropy (Kullback-Leibler distance) between the prior distribution Pr(C) and the posterior distribution Pr(C | w).

$$D(\Pr(C|w) \parallel \Pr(C)) = \sum_c \Pr(c|w) \log \frac{\Pr(c|w)}{\Pr(c)} \quad (2)$$

$$= \sum_c \Pr(c|w) \, I(c; w) \quad (3)$$

Here w is a word with selectional properties, C ranges over semantic classes, and co-occurrences are counted with respect to a particular argument — e.g. verbs and direct objects, nominal modifiers and the head noun they modify, and so forth. Intuitively, this definition works by comparing the distribution of argument classes *without* knowing what the word is (e.g., the *a priori* likelihood of classes in direct object position), to the distribution *with respect to* the word. If these distributions are very different, as measured by relative entropy, then the word has a strong influence on what can or cannot appear in that argument position, and we say that it has a strong selectional preference for that argument.

The "goodness of fit" between a word and a particular class of arguments is captured by the following definition:

Definition. The *selectional association* of w with c is the contribution c makes to the selectional preference of w.

$$A(w, c) = \frac{\Pr(c|w) \log \frac{\Pr(c|w)}{\Pr(c)}}{D(\Pr(C|w) \parallel \Pr(C))} \quad (4)$$

The selectional association $A(w_1, w_2)$ of two words is taken to be the maximum of $A(w_1, c)$ over all classes c to which w_2 belongs.

VERB, ARGUMENT	"BEST" ARGUMENT CLASS	A
drink wine	⟨beverage⟩	0.088
drink gasoline	⟨substance⟩	0.075
drink pencil	⟨object⟩	0.030
drink sadness	⟨psychological_feature⟩	-0.001

The above table illustrates how this definition captures the qualitative differences in example (2). The "best" class for an argument is the class that maximizes selectional association. Notice that finding that class represents a form of sense disambiguation using local context (cf. [15]): of all the classes to which the noun *wine* belongs — including ⟨alcohol⟩, ⟨substance⟩, ⟨red⟩, and ⟨color⟩, among others — the class ⟨beverage⟩ is the sense of *wine* most appropriate as a direct object for *drink*.

3.2. Semantic Similarity

Any number of factors influence judgements of semantic similarity between two nouns. Here we propose to use only one source of information: the relationship between classes in the WordNet IS-A taxonomy. Intuitively, two noun classes can be considered similar when there is a single, specific class that subsumes them both — if you have to travel very high in the taxonomy to find a class that subsumes both classes, in the extreme case all the way to the top, then they cannot have all that much in common. For example, ⟨nickel⟩ and ⟨dime⟩ are both immediately subsumed by ⟨coin⟩, whereas the most specific superclass that ⟨nickel⟩ and ⟨mortgage⟩ share is ⟨possession⟩.

The difficulty, of course, is how to determine which superclass is "most specific." Simply counting IS-A links in the taxonomy can be misleading, since a single link can represent a fine-grained distinction in one part of the taxonomy (e.g. ⟨zebra⟩ IS-A ⟨equine⟩) and a very large distinction elsewhere (e.g. ⟨carcinogen⟩ IS-A ⟨substance⟩).

Rather than counting links, we use the *information content* of a class to measure its specificity (i.e., $-\log \Pr(c)$); this permits us to define noun similarity as follows:

Definition. The *semantic similarity* of n_1 and n_2 is

$$\text{sim}(n_1, n_2) = \sum_i \alpha_i [-\log \Pr(c_i)], \quad (5)$$

where $\{c_i\}$ is the set of classes dominating both n_1 and n_2. The α_i, which sum to 1, are used to weight the contribution of each class — for example, in accordance with word sense probabilities. In the absence of word sense constraints we can compute the "globally" most specific class simply by setting α_i to 1 for the class maximizing $[-\log \Pr(c)]$,

279

and 0 otherwise. For example, according to that "global" measure, sim(*nickel,dime*) = 12.71 (= − log Pr(⟨*coin*⟩)) and sim(*nickel,mortgage*) = 7.61 (= − log Pr(⟨*possession*⟩)).

4. SYNTACTIC AMBIGUITY

4.1. Coordination and Nominal Compounds

Having proposed formalizations of selectional preference and semantic similarity as information-theoretic relationships involving conceptual classes, we now turn to the application of these ideas to the resolution of syntactic ambiguity.

Ambiguous coordination is a common source of parsing difficulty. In this study, we investigated the application of class-based statistical methods to a particular subset of coordinations, noun phrase conjunctions of the form *noun1 and noun2 noun3*, as in (3):

(3)   a. a (bank and warehouse) guard
       b. a (policeman) and (park guard)

Such structures admit two analyses, one in which *noun1* and *noun2* are the two heads being conjoined (3a) and one in which the conjoined heads are *noun1* and *noun3* (3b).

As pointed out by Kurohashi and Nagao [7], similarity of form and similarity of meaning are important cues to conjoinability. In English, similarity of form is to a great extent captured by agreement in number:

(4)   a. several *business* and *university* groups
       b. several *businesses* and university *groups*

Semantic similarity of the conjoined heads also appears to play an important role:

(5)   a. a *television* and *radio* personality
       b. a *psychologist* and sex *researcher*

In addition, for this particular construction, the appropriateness of noun-noun modification for *noun1* and *noun3* is relevant:

(6)   a. *mail* and securities *fraud*
       b. *corn* and peanut *butter*

We investigated the roles of these cues by conducting a disambiguation experiment using the definitions in the previous section. Two sets of 100 noun phrases of the form [NP *noun1 and noun2 noun3*] were extracted from the *Wall Street Journal* (WSJ) corpus in the Penn Treebank and disambiguated by hand, with one set to be used for development and the other for testing.[3] A set of simple transformations were applied to all WSJ data, including the mapping of all

---

[3]Hand disambiguation was necessary because the Penn Treebank does not encode NP-internal structure. These phrases were disambiguated using the full sentence in which they occurred, plus the previous and following sentence, as context.

proper names to the token *someone*, the expansion of month abbreviations, and the reduction of all nouns to their root forms.

Similarity of form, defined as agreement of number, was determined using a simple analysis of suffixes in combination with WordNet's database of nouns and noun exceptions. Similarity of meaning was determined "globally" as in equation (5) and the example that followed; noun class probabilities were estimated using a sample of approximately 800,000 noun occurrences in Associated Press newswire stories.[4] For the purpose of determining semantic similarity, nouns not in WordNet were treated as instances of the class ⟨thing⟩. Appropriateness of noun-noun modification was determined using selectional association as defined in equation (4), with co-occurrence frequencies calculated using a sample of approximately 15,000 noun-noun compounds extracted from the WSJ corpus. (This sample did not include the test data.) Both selection of the modifier for the head and selection of the head for the modifier were considered.

Each of the three sources of information — form similarity, meaning similarity, and modification relationships — was used alone as a disambiguation strategy, as follows:

- Form:

   - If noun1 and noun2 match in number
     and noun1 and noun3 do not
     then conjoin noun1 and noun2;

   - if noun1 and noun3 match in number
     and noun1 and noun2 do not
     then conjoin noun1 and noun3;

   - otherwise remain undecided.

- Meaning:

   - If sim(noun1,noun2) > sim(noun1,noun3)
     then conjoin noun1 and noun2;

   - if sim(noun1,noun3) > sim(noun1,noun2)
     then conjoin noun1 and noun3;

   - otherwise remain undecided.

- Modification:

   - If A(noun1,noun3) > τ, a threshold, *or*
     if A(noun3,noun1) > τ,
     then conjoin noun1 and noun3;

   - If A(noun1,noun3) < σ *and* A(noun3,noun1) < σ
     then conjoin noun1 and noun2;

   - otherwise remain undecided.[5]

In addition, we investigated several methods for combining the three sources of information. These included: (a) "backing off" (i.e., given the form, modification, and meaning

---

[4]I am grateful to Donald Hindle for making these data available.
[5]Thresholds τ and σ were fixed before evaluating the test data.

strategies in that order, use the first strategy that isn't undecided); (b) taking a "vote" among the three strategies and choosing the majority; (c) classifying using the results of a linear regression; and (d) constructing a decision tree classifier.

The training set contained a bias in favor of conjoining *noun1* and *noun2*, so a "default" strategy — always choosing that bracketing — was used as a baseline. The results are as follows:

STRATEGY	ANSWERED (%)	PRECISION (%)
Default	100.0	66.0
Form	53.0	90.6
Modification	75.0	69.3
Meaning	66.0	71.2
Backing off	95.0	81.1
Voting	89.0	78.7
Regression	100.0	79.0
ID3 Tree	100.0	80.0

Not surprisingly, the individual strategies perform reasonably well on the instances they can classify, but recall is poor; the strategy based on similarity of form is highly accurate, but arrives at an answer only half the time. Of the combined strategies, the "backing off" approach succeeds in answering 95% of the time and achieving 81.1% precision — a reduction of 44.4% in the baseline error rate.

We have recently begun to investigate the disambiguation of more complex coordinations of the form [NP *noun1 noun2 and noun3 noun4*], which permit five possible bracketings:

(7)   a.   freshman ((business and marketing) major)
       b.   (food (handling and storage)) procedures
       c.   ((mail fraud) and bribery) charges

       d.   Clorets (gum and (breath mints))
       e.   (baby food) and (puppy chow)

These bracketings comprise two groups, those that conjoin *noun2* and *noun3* (a–c) and those that conjoin *noun2* and *noun4* (d–e). Rather than tackling the five-way disambiguation problem immediately, we began with an experimental task of classifying a noun phrase as belonging to one of these two groups.

We examined three classification strategies. First, we used the form-based strategy described above. Second, as before, we used a strategy based on semantic similarity; this time, however, selectional association was used to determine the α_i in equation (5), incorporating modifier-head relationships into the semantic similarity strategy. Third, we used "backing off" (from form similarity to semantic similarity) to combine the two individual strategies. As before, one set of items was used

for development, and another set (89 items) was set aside for testing. As a baseline, results were evaluated against a simple default strategy of always choosing the group that was more common in the development set.

STRATEGY	ANSWERED (%)	PRECISION (%)
Default	100.0	44.9
Form	40.4	80.6
Meaning	69.7	77.4
Backing off	85.4	81.6

In this case, the default strategy defined using the development set was misleading, leading to worse than chance precision. However, even if default choices were made using the bias found in the test set, precision would be only 55.1%. The results in the above table make it clear that the strategies using form and meaning are far more accurate, and that combining them leads to good coverage and precision.

The pattern of results in these two experiments demonstrates a significant reduction in syntactic misanalyses for this construction as compared to the simple baseline, and it confirms that form, meaning, and modification relationships all play a role in disambiguation. In addition, these results confirm the effectiveness of the proposed definitions of selectional preference and semantic similarity.

4.2. Prepositional Phrase Attachment[6]

Prepositional phrase attachment represents another important form of parsing ambiguity. Empirical investigation [14] suggests that lexical preferences play an important role in disambiguation, and Hindle and Rooth [5] have demonstrated that these preferences can be acquired and utilized using lexical co-occurrence statistics.

(8)   a.   They foresee little progress in exports.
       b.   [VP foresee [NP little progress [PP in exports]]]
       c.   [VP foresee [NP little progress] [PP in exports]]

Given an example such as (8a), Hindle and Rooth's "lexical association" strategy chooses between bracketings (8b) and (8c) by comparing $Pr(in|foresee)$ with $Pr(in|progress)$ and evaluating the direction and significance of the difference between the two conditional probabilities. The object of the preposition is ignored, presumably because the data would be far too sparse if it were included.

As Hearst and Church [4] observe, however, the object of the preposition can provide crucial information for determining attachment, as illustrated in (9):

(9)   a.   Britain reopened its embassy in December.
       b.   Britain reopened its embassy in Teheran.

---

[6]This section reports work done in collaboration with Marti A. Hearst.

Hoping to overcome the sparseness problem and use this information, we formulated a strategy of "conceptual association," according to which the objects of the verb and preposition are treated as members of semantic classes and the two potential attachment sites are evaluated using class-based rather than lexical statistics.

The alternative attachment sites — verb-attachment and noun-attachment — were evaluated according to the following criteria:

$$\text{vscore} = \text{freq(v,PP) I(v;PP)} \quad (6)$$
$$\text{nscore} = \text{freq(class1,PP) I(class1;PP)} \quad (7)$$

where PP is an abbreviation for (preposition,class2), and class1 and class2 are classes to which the object of the verb and object of the preposition belong, respectively. These scores were used rather than conditional probabilities Pr(PP | v) and Pr(PP | class1) because, given a set of possible classes to use as class2 (e.g. *export* is a member of ⟨export⟩, ⟨commerce⟩, ⟨group_action⟩, and ⟨human_action⟩), conditional probability will always favor the most general class. In contrast, comparing equations (6) and (7) with equation (4), the verb- and noun-attachment scores resemble the selectional association of the verb and noun with the prepositional phrase.

Because nouns belong to many classes, we required some way to combine scores obtained under different classifications. Rather than considering the entire cross-product of classifications for the object of the verb and the object of the preposition, we chose to first consider all possible classifications of the object of the preposition, and then to classify the object of the verb by choosing class1 so as to maximize I(class1;PP). For example, sentence (8a) yields the following classifications:

CLASS1	PP	NSCORE	VSCORE
⟨situation⟩	in ⟨export⟩	67.4	39.8
⟨rise⟩	in ⟨commerce⟩	178.3	23.8
⟨advance⟩	in ⟨group_action⟩	104.9	19.9
⟨advance⟩	in ⟨act⟩	149.5	40.6

The "conceptual association" strategy merges evidence from alternative classifications in an extremely simple way: by performing a paired samples t-test on the nscores and vscores, and preferring attachment to the noun if t is positive, and to the verb if negative. A combined strategy uses this preference if t is significant at $p < .1$, and otherwise uses the lexical association preference. For example (8a), $t(3) = 3.57, p < .05$, with (8b) being the resulting choice of bracketing.

We evaluated this technique using the Penn Treebank *Wall Street Journal* corpus, comparing the performance of lexical association alone (LA), conceptual association alone (CA), and the combined strategy (COMBINED) on a held-out set of 174 ambiguous cases. The results were as follows:

	LA	CA	COMBINED
% Correct	81.6	77.6	82.2

When the individual strategies were constrained to answer only when confident ($|t| > 2.1$ for lexical association, $p < .1$ for conceptual association), they performed as follows:

STRATEGY	Answered (%)	Precision (%)
LA	44.3	92.8
CA	67.2	84.6

Despite the fact that this experiment used an order of magnitude less training data than Hindle and Rooth's, their lexical association strategy performed quite a bit better than in the experiments reported in [5], presumably because this experiment used hand-disambiguated rather than heuristically disambiguated training data.

In this experiment, the bottom-line performance of the conceptual association strategy is worse than that of lexical association, and the combined strategy yields at best a marginal improvement. However, several observations are in order. First, the coverage and precision achieved by conceptual association demonstrate some utility of class information, since the lexical data are impossibly sparse when the object of the preposition is included. Second, a qualitative evaluation of what conceptual association actually did shows that it is capturing relevant relationships for disambiguation.

(10)   To keep his schedule on track, he flies two personal secretaries in from Little Rock to *augment his staff in Dallas.*

For example, *augment* and *in* never co-occur in the training corpus, and neither do *staff* and *in*; as a result, the lexical association strategy makes an incorrect choice for the ambiguous verb phrase in (10). However, the conceptual association strategy makes the correct choice on the basis of the following classifications:

CLASS1	PP	NSCORE	VSCORE
⟨gathering⟩	in ⟨dallas⟩	38.18	45.54
⟨people⟩	in ⟨urban_area⟩	1200.21	28.46
⟨personnel⟩	in ⟨region⟩	314.62	23.38
⟨personnel⟩	in ⟨geographical_area⟩	106.05	26.80
⟨people⟩	in ⟨city⟩	1161.22	28.61
⟨personnel⟩	in ⟨location⟩	320.85	22.83

Third, mutual information appears to be a successful way to select appropriate classifications for the direct object, given a classification of the object of the preposition. For example, despite the fact that *staff* belongs to 25 classes in WordNet — including ⟨musical_notation⟩ and ⟨rod⟩, for instance — the classes to which it is assigned in the above table seem contextually appropriate. Finally, it is clear that in many instances

the paired t-test, which effectively takes an unweighted average over multiple classifications, is a poor way to combine sources of evidence.

In two additional experiments, we examined the effect of semantic classes on robustness, since presumably a domain-independent source of noun classes should be able to mitigate the effects of a mismatch between training data and test data. In the first of these experiments, we used the WSJ training material, and tested on 173 instances from Associated Press newswire, with the following results:

	LA	CA	COMBINED
% Correct	69.9	72.3	72.8

STRATEGY	ANSWERED (%)	PRECISION (%)
LA	31.8	80.0
CA	49.7	77.9

In the second experiment, we retained the *test* material from the WSJ corpus, but trained on the Brown corpus material in the Penn Treebank. The results were as follows:

	LA	CA	COMBINED
% Correct	77.6	73.6	79.3

STRATEGY	ANSWERED (%)	PRECISION (%)
LA	35.6	85.5
CA	59.2	81.6

These additional experiments demonstrate large increases in coverage when confident (55–65%) with only moderate decreases in precision ($< 5\%$). Overall, the results of the three experiments seem promising, and suggest that further work on conceptual association will yield improvements to disambiguation strategies using lexical association alone.

5. Conclusions

In this paper, we have used a knowledge-based conceptual taxonomy, together with corpus-based lexical statistics, to provide new formalizations of selectional preference and semantic similarity. Although a complete characterization of these and other semantic notions may ultimately turn out to require a full-fledged theory of meaning, lexical-conceptual representation, and inference, we hope to have shown that a great deal can be accomplished using a simple semantic representation combined with appropriate information-theoretic ideas. Conversely, we also hope to have shown the utility of knowledge-based semantic classes in arriving at a statistical characterization of linguistic phenomena, as compared to purely distributional methods. A detailed comparison of knowledge-based and distributionally-derived word classes is

needed in order to assess the advantages and disadvantages of each approach.

"Every way ambiguous" constructions form a natural class of practical problems to investigate using class-based statistical techniques. The present results are promising, and we are exploring improvements to the particular algorithms and results illustrated here. In future work we hope to investigate other ambiguous constructions, and to explore the implications of selectional preference for word-sense disambiguation.

References

1. Brown, P., V. Della Pietra, P. deSouza, J. Lai, and R. Mercer, "Class-based N-gram Models of Natural Language," Computational Linguistics 18(4), December, 1992.
2. Church, K. W. and R. Patil, "Coping with Syntactic Ambiguity or How to Put the Block in the Box on the Table," American Journal of Computational Linguistics, 8(3-4), 1982.
3. Good, I.J., "The Population Frequencies of Species and the Estimation of Population Parameters," Biometrika 40(3 and 4), pp. 237-264, (1953).
4. Hearst, M. A. and K. W. Church, "An Investigation of the Use of Lexical Associations for Prepositional Phrase Attachment," in preparation.
5. Hindle, D., "Noun Classification from Predicate-Argument Structures," *Proceedings of the 28th Annual Meeting of the Assocation of Computational Linguistics*, 1990.
6. Hindle, D. and M. Rooth, "Structural Ambiguity and Lexical Relations," *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, 1991.
7. Kurohashi, S. and M. Nagao, "Dynamic Programming Method for Analyzing Conjunctive Structures in Japanese," Proceedings of COLING–92, Nantes, France, August, 1992.
8. Miller, G. A., "WordNet: An On-Line Lexical Database," International Journal of Lexicography 3(4), 1990.
9. Pereira, Fernando and Naftali Tishby, "Distributional Similarity, Phase Transitions and Hierarchical Clustering," presented at the AAAI Fall Symposium on Probabilistic Approaches to Natural Language, Cambridge, Massachusetts, October, 1992.
10. Resnik, Philip, "WordNet and Distributional Analysis: A Class-based Approach to Lexical Discovery," AAAI Workshop on Statistically-based NLP Techniques, San Jose, California, July, 1992.
11. Resnik, Philip, "Selectional Preference and Implicit Objects," CUNY Sentence Processing Conference, Amherst, Massachusetts, March, 1993.
12. Schuetze, Hinrich, "Word Space," in Hanson, S. J., J. D. Cowan, and C. L. Giles (eds.) *Advances in Neural Information Processing Systems 5*, Morgan Kaufmann, to appear.
13. Weischedel, R., M. Meteer, R. Schwartz, and J. Palmucci, "Coping with Ambiguity and Unknown Words through Probabilistic Models," DARPA workshop, 1989.
14. Whittemore, G., K. Ferrara,and H. Bruner, "Empirical Study of Predictive Powers of Simple AttachmentSchemes for Postmodifier Prepositional Phrases," *Proceedings of the 28th Annual Meeting of the Assocation of Computational Linguistics*, 1990.
15. Yarowsky, David, "One Sense Per Collocation," this volume.
16. Zernik, Uri, ed., *Lexical Acquisition: Using On-line Resources to Build a Lexicon*, Lawrence Erlbaum, 1991.