

A CLASS-BASED APPROACH TO LEXICAL DISCOVERY

Philip Resnik*

Department of Computer and Information Science, University of Pennsylvania
Philadelphia, Pennsylvania 19104, USA

Internet: *resnik@linc.cis.upenn.edu*

1 Introduction

In this paper I propose a generalization of lexical association techniques that is intended to facilitate statistical discovery of facts involving word *classes* rather than individual words. Although defining association measures over classes (as sets of words) is straightforward in theory, making direct use of such a definition is impractical because there are simply too many classes to consider. Rather than considering all possible classes, I propose constraining the set of possible word classes by using a broad-coverage lexical/conceptual hierarchy [Miller, 1990].

2 Word/Word Relationships

Mutual information is an information-theoretic measure of association frequently used with natural language data to gauge the “relatedness” between two words x and y . It is defined as follows:

$$I(x; y) = \log \frac{\Pr(x, y)}{\Pr(x)\Pr(y)} \quad (1)$$

As an example of its use, consider Hindle’s [1990] application of mutual information to the discovery of predicate argument relations. Hindle investigates word co-occurrences as mediated by syntactic structure. A six-million-word sample of Associated Press news stories was parsed in order to construct a collection of subject/verb/object instances. On the basis of these data, Hindle calculates a *co-occurrence score* (an estimate of mutual information) for verb/object pairs and verb/subject pairs. Table 1 shows some of the verb/object pairs for the verb *drink* that occurred more than once, ranked by co-occurrence score, “in effect giving the answer to the question ‘what can you drink?’ ” [Hindle, 1990], p. 270.

Word/word relationships have proven useful, but are not appropriate for all applications. For example,

*This work was supported by the following grants: ARO DAAL 03-89-C-0031, DARPA N00014-90-J-1863, NSF IRI 90-16592, Ben Franklin 91S.3078C-1. I am indebted to Eric Brill, Henry Gleitman, Lila Gleitman, Aravind Joshi, Christine Nakatani, and Michael Niv for helpful discussions, and to George Miller and colleagues for making WordNet available.

Co-occurrence score	verb	object
11.75	drink	tea
11.75	drink	Pepsi
11.75	drink	champagne
10.53	drink	liquid
10.20	drink	beer
9.34	drink	wine

Table 1: High-scoring verb/object pairs for *drink* (part of Hindle 1990, Table 2).

the selectional preferences of a verb constitute a relationship between a verb and a class of nouns rather than an individual noun.

3 Word/Class Relationships

3.1 A Measure of Association

In this section, I propose a method for discovering *class-based* relationships in text corpora on the basis of mutual information, using for illustration the problem of finding “prototypical” object classes for verbs.

Let $\mathcal{V} = \{v_1, v_2, \dots, v_l\}$ and $\mathcal{N} = \{n_1, n_2, \dots, n_m\}$ be the sets of verbs and nouns in a vocabulary, and $\mathcal{C} = \{c | c \subseteq \mathcal{N}\}$ the set of noun classes; that is, the power set of \mathcal{N} . Since the relationship being investigated holds between verbs and classes of their objects, the elementary events of interest are members of $\mathcal{V} \times \mathcal{C}$. The joint probability of a verb and a class is estimated as

$$\Pr(v, c) \approx \frac{\sum_{n \in c} \text{count}(v, n)}{\sum_{v' \in \mathcal{V}} \sum_{n' \in \mathcal{N}} \text{count}(v', n')} \quad (2)$$

Given $v \in \mathcal{V}$, $c \in \mathcal{C}$, define the *association score*

$$A(v, c) \triangleq \Pr(c|v) \log \frac{\Pr(v, c)}{\Pr(v)\Pr(c)} \quad (3)$$

$$= \Pr(c|v)I(v; c). \quad (4)$$

The association score takes the mutual information between the verb and a class, and scales it according

to the likelihood that a member of that class will actually appear as the object of the verb.¹

3.2 Coherent Classes

A search among a verb’s object nouns requires at most $|\mathcal{N}|$ computations of the association score, and can thus be done exhaustively. An exhaustive search among object *classes* is impractical, however, since the number of classes is exponential. Clearly some way to constrain the search is needed. I propose restricting the search by imposing a requirement of *coherence* upon the classes to be considered. For example, among possible classes of objects for *open*, the class $\{\textit{closet, locker, store}\}$ is more coherent than $\{\textit{closet, locker, discourse}\}$ on intuitive grounds: every noun in the former class describes a repository of some kind, whereas the latter class has no such obvious interpretation.

The WordNet lexical database [Miller, 1990] provides one way to structure the space of noun classes, in order to make the search computationally feasible. WordNet is a lexical/conceptual database constructed on psycholinguistic principles by George Miller and colleagues at Princeton University. Although I cannot judge how well WordNet fares with regard to its psycholinguistic aims, its noun taxonomy appears to have many of the qualities needed if it is to provide basic taxonomic knowledge for the purpose of corpus-based research in English, including broad coverage and multiple word senses.

Given the WordNet noun hierarchy, the definition of “coherent class” adopted here is straightforward. Let $\textit{words}(w)$ be the set of nouns associated with a WordNet class w .²

Definition. A noun class $c \in \mathcal{C}$ is *coherent* iff there is a WordNet class w such that $\textit{words}(w) \cap \mathcal{N} = c$.

As a consequence of this definition, noun classes that are “too small” or “too large” to be coherent are excluded, and the problem of search through an exponentially large space of classes is reduced to search within the WordNet hierarchy.³

¹Scaling mutual information in this fashion is often done; see, e.g., [Rosenfeld and Huang, 1992].

²Strictly speaking, WordNet as described by [Miller, 1990] does not have classes, but rather lexical groupings called synonym sets. By “WordNet class” I mean a pair (word, synonym-set).

³A related possibility being investigated independently by Paul Kogut (personal communication) is assign to each noun and verb a vector of feature/value pairs based upon the word’s classification in the WordNet hierarchy, and to classify nouns on the basis of their feature-value correspondences.

A(v,c)	verb	object class
3.58	drink	$\langle \textit{beverage, [beverage, \dots]} \rangle$
2.05	drink	$\langle \textit{intoxicant, [alcohol, \dots]} \rangle$

Table 2: Object classes for *drink*

4 Preliminary Results

An experiment was performed in order to discover the “prototypical” object classes for a set of 115 common English verbs. The counts of equation (2) were calculated by collecting a sample of verb/object pairs from the Brown corpus.⁴ Direct objects were identified using a set of heuristics to extract only the *surface object* of the verb. Verb inflections were mapped down to the base form and plural nouns mapped down to singular.⁵ For example, the sentence *John ate two shiny red apples* would yield the pair (*eat, apple*). The sentence *These are the apples that John ate* would not provide a pair for *eat*, since *apple* does not appear as its surface object.

Given each verb, v , the “prototypical” object class was found by conducting a best-first search upwards in the WordNet noun hierarchy, starting with WordNet classes containing members that appeared as objects of the verb. Each WordNet class w considered was evaluated by calculating $A(v, \{n \in \mathcal{N} | n \in \textit{words}(w)\})$. Classes having too low a count (fewer than five occurrences with the verb) were excluded from consideration.

The results of this experiment are encouraging. Table 2 shows the object classes discovered for the verb *drink* (compare to Table 1), and Table 3 the highest-scoring object classes for several other verbs. Recall from the definition in Section 3.2 that each WordNet class w in the tables appears as an abbreviation for $\{n \in \mathcal{N} | n \in \textit{words}(w)\}$; for example, $\langle \textit{intoxicant, [alcohol, \dots]} \rangle$ appears as an abbreviation for $\{\textit{whisky, cognac, wine, beer}\}$.

5 Acquisition of Verb Properties

More work is needed to improve the performance of the technique proposed here. At the same time, the ability to approximate a lexical/conceptual classification of nouns opens up a number of possible applications in lexical acquisition. What such applications have in common is the use of lexical associations as a window into semantic relationships. The technique described in this paper provides a new, hierarchical

⁴The version of the Brown corpus used was the tagged corpus found as part of the Penn Treebank.

⁵Nouns outside the scope of WordNet that were tagged as proper names were mapped to the token *pname*, a subclass of classes $\langle \textit{someone, [person]} \rangle$ and $\langle \textit{location, [location]} \rangle$.

A(v,c)	verb	object class
1.94	ask	{question, [question, ...]}
0.16	call	{someone, [person, ...]}
2.39	climb	{stair, [step, ...]}
3.64	cook	{repast, [repast, ...]}
0.27	draw	{cord, [cord, ...]}
3.58	drink	{beverage, [beverage, ...]}
1.76	eat	{nutrient, [food, ...]}
0.30	lose	{sensory_faculty, [sense, ...]}
1.28	play	{part, [character, ...]}
2.48	pour	{liquid, [liquid, ...]}
1.03	pull	{cover, [covering, ...]}
1.23	push	{button, [button, ...]}
1.18	read	{written_material, [writing, ...]}
2.69	sing	{music, [music, ...]}

Table 3: Some "prototypical" object classes

source of semantic knowledge for statistical applications. This section briefly discusses one area where this kind of knowledge might be exploited.

Diathesis alternations are variations in the way that a verb syntactically expresses its arguments [Levin, 1989]. For example, 1(a,b) shows an instance of the *indefinite object alternation*, and 2(a,b) shows an instance of the *causative/inchoative alternation*.

- 1
 - a. John ate lunch.
 - b. John ate.
- 2
 - a. John opened the door.
 - b. The door opened.

Such phenomena are of particular interest in the study of how children learn the semantic and syntactic properties of verbs, because they stand at the border of syntax and lexical semantics. There are numerous possible explanations for why verbs fall into particular classes of alternations, ranging from shared semantic properties of verbs within a class, to pragmatic factors, to "lexical idiosyncrasy."

Statistical techniques like the one described in this paper may be useful in investigating relationships between verbs and their arguments, with the goal of contributing data to the study of diathesis alternations, and, ideally, in constructing a computational model of verb acquisition. For example, in the experiment described in Section 4, the verbs participating in "implicit object" alternations⁶ appear to have higher association scores with their "prototypical" object classes than verbs for which implicit objects are disallowed. Preliminary results, in fact, show a statistically significant difference between the two groups.

⁶The indefinite object alternation [Levin, 1989] and the specified object alternation [Cote, 1992].

Might such shared information-theoretic properties of verbs play a role in their acquisition, in the same way that shared semantic properties might?

On a related topic, Grimshaw has recently suggested that the syntactic bootstrapping hypothesis for verb acquisition [Gleitman, 1991] be extended in such a way that alternations such as the causative/inchoative alternation (e.g. 2(a,b)) are learned using class information about the observed subjects and objects of the verb, in addition to subcategorization information.⁷ I hope to extend the work on verb/object associations described here to other arguments of the verb in order to explore this suggestion.

6 Conclusions

The technique proposed here provides a way to study statistical associations beyond the level of individual words, using a broad-coverage lexical/conceptual hierarchy to structure the space of possible noun classes. Preliminary results, on the task of discovering "prototypical" object classes for a set of common English verbs, appear encouraging, and applications in the study of verb argument structure are apparent. In addition, assuming that the WordNet hierarchy (or some similar knowledge base) proves appropriately broad and consistent, the approach proposed here may provide a model for importing basic taxonomic knowledge into other corpus-based investigations, ranging from computational lexicography to statistical language modelling.

References

- [Cote, 1992] Sharon Cote. Discourse functions of two types of null objects in English. Presented at the 66th Annual Meeting of the Linguistic Society of America, Philadelphia, PA, January 1992.
- [Gleitman, 1991] Lila Gleitman. The structural sources of verb meanings. *Language Acquisition*, 1, 1991.
- [Hindle, 1990] Donald Hindle. Noun classification from predicate-argument structures. In *Proceedings of the 28th Annual Meeting of the ACL*, 1990.
- [Levin, 1989] Beth Levin. Towards a lexical organization of English verbs. Technical report, Dept. of Linguistics, Northwestern University, November 1989.
- [Miller, 1990] George Miller. Wordnet: An on-line lexical database. *International Journal of Lexicography*, 4(3), 1990. (Special Issue).
- [Rosenfeld and Huang, 1992] Ronald Rosenfeld and Xuedong Huang. Improvements in stochastic language modelling. In Mitch Marcus, editor, *Fifth DARPA Workshop on Speech and Natural Language*, February 1992. Arden House Conference Center, Harriman, NY.

⁷Jane Grimshaw, keynote address, Lexicon Acquisition Workshop, University of Pennsylvania, January, 1992.