Lexical Substitutability

1. Too Much and Too Little Evidence

In their attempts to describe the contemporary language, lexicographers are both overwhelmed by evidence and at the same time starved of evidence. Everything they ever read or hear or say is evidence for the existence of (or even for an answer to) some lexicographic question, but most of it flashes by so fast that it is impossible to capture. Even in these days of large electronic corpora, it is often not possible for a lexicographer to find precisely the right little bit of evidence that he or she needs at precisely the moment it is needed, to provide the answer to some question of detail. Indeed, we cannot always be sure that we are asking the right questions. A lexicographer is like a person standing underneath Niagara Falls holding a rainwater gauge, while the evidence sweeps by in immeasurable torrents.

Lexicographers differ in their responses to this desperate situation. Some seek to capture particularly fine and unusual droplets of spray: they collect citations for rare words and uses, while ignoring the torrent of ordinary, everyday usage that is sweeping past them. Such eclecticism would be more acceptable if we could be sure that ordinary, everyday usage is fully and accurately described in current dictionaries. A statistician would be concerned that a human reading program for a citation collection might not be the most appropriate sampling procedure, especially if one wants to describe common usage.

Others turn inwards, and report their own intuitions. This, like citation collection, can be a valuable and informative exercise as far as it goes, but it has nothing to say about any patterns of linguistic behavior that might be observable in the language use of the community at large.

Yet others, by the ingenious device of an electronic corpus, have managed to freeze a small part of the torrent in order to climb all over it. Very large corpora are beginning to be widely available, but techniques for corpus analysis are still in their infancy. The basic analytic tool is a set of concordances for each linguistic type in the corpus: the lexicographer tries to group the concordance lines into meanings and usage patterns. As corpora get larger, this task becomes increasingly daunting. Even a 20-million word corpus contains hundreds of occurrences of common words such as *ask* and *help*, and these can take days and even weeks to analyse, not always with satisfactory results: the unaided human mind can be simply overwhelmed by data. And lexicography is now entering an era in which corpora of over 100 million

words will be commonplace.

Over the last few years, we have been developing a number of tools which we hope will help lexicographers to analyse evidence from large corpora, for example by providing various statistical summaries of the evidence. These tools enhance the value of raw concordances by refining the massive volumes of evidence down to a more manageable amount of relatively high-grade lexicographic ore. In previous work, (Church and Hanks, 1990) and (Church, Gale, Hanks, Hindle, 1990), we have discussed a number of well-known statistical tools that have been used successfully in other applications such as speech recognition and machine translation (Jelinek, 1985). In particular, we have found mutual information to be a useful measure of similarity, and the *t*-score to be a useful measure of difference. Both tools are useful for summarizing the concordance evidence in ways that make it more manageable and more accessible, although considerable human judgement (and hard work) is still required, for example in order to decide what to compare and contrast with what.

In this paper we propose a further tool, a measure of substitutability, henceforth the *sub*-test, which is similar in many respects to the independently developed method of Justeson and Katz (1991). The *sub*-test identifies sets of words that have similar distributions, for example *request* and *ask for*. We find that substitutable words are often close synonyms (or antonyms), although, as we shall see, other kinds of relations can also be observed. We hope that this will help lexicographers and thesaurus writers to decide which sets and pairs of words they should compare and contrast on the basis of corpus evidence.

The idea of using substitution as a basis for synonymy has a long history, dating back to Leibniz's famous definition of identity (1704): *eadem sunt quorum unum potest substitui alteri salva veritate* (''Two things are identical if one can be substituted for the other without affecting the truth''). Quine (1960) clarified a number of possible confusions about Leibniz's definition and showed its relevance to language. He argued that the two things in question are signs (for example symbols, words, or propositions), rather than the objects in the world denoted by these signs. Ullmann (1962: 141ff.), Lyons (1977: 197ff.), and others have followed this tradition in their definitions of synonymy.

2. Categorical Synonymy vs. Gradient Synonymy

Although it is natural to think of synonymy as an equivalence relation or a partial order (e.g., an IS-A hierarchy), we have specifically designed the *sub*-test so that it is neither symmetric nor transitive. Few pairs of words fit very well into an equivalence relation: the study of lexis shows that where the conditions have been met, more or less, in the past, there is attrition of one form or specialization of one form. Consider, for example, such pairs as *attic/garret, smallpox/variola, royal/regal, page/leaf.* The same phenomenon can be observed in other languages: Ullmann refers to cases in French where pairs such as *rivière* and *fleuve* have been officially distinguished (1962: 145). This encourages the view that there is no such thing as categorical or absolute synonymy.

Ron Hardin (personal communication) has come to the same basic conclusion after running some path finding software on a version of *The New Collins Thesaurus* (McLeod, 1984) that has been processed in certain ways that we don't completely understand. He found that you can go from any word (e.g., *authentic*) to its antonym (*unauthentic*) in just a few steps: *authentic* → *believable* → *probable* → *ostensible* → *pretended* → *spurious* → *unauthentic*. Each of the steps is reasonable enough, but after composing enough of them, you have gone from hot to cold. There are rarely more than six steps between any synonym/antonym pair.

The categorical view of synonymy also fails to account for certain observations about spoken discourse, which suggest that speakers constantly negotiate meanings. It is important to allow for gradient synonymy in order to understand a discourse such as the following:

A: Bad, isn't it?
B: Yes, it's shocking.

It might have been thought that B's use of a different word, *shocking* as opposed to *bad*, indicates disagreement with A. In fact, near synonyms and antonyms contribute to the lexical cohesion of a text, as observed by Halliday and Hasan (1976: 277 ff.) and Brazil (1985: passim).

For these reasons, we prefer to view synonymy as a gradient property rather than an all-or-nothing categorical relation, as an equivalence relation or a partial order would have it. This view of graded

category membership is similar in spirit to radial categories as described by Lakoff (1987). Cruse (1986: 265-291) also makes use of a gradient view of synonymy (which he calls *intuitive* synonymy), and distinguishes this view from a categorical view (which he calls *cognitive* or *purist* synonymy).

3. What Counts as Evidence for Synonymy in Reference Works?

3.1 Little Agreement Across Dictionaries and Thesauruses

It is hard to know when two words should be considered synonymous. The synonym studies in the big American unabridged dictionaries illustrate the problem. It is remarkable just how little agreement there is among the various reference works.

Consider, for example, the synonym studies for *ask* in the *Random House Dictionary, Second Edition* (Flexner et al., 1987), henceforth RHD2, and those in *Merriam Webster's Third New International* (Gove et al., 1961), henceforth MW3. RHD2 discusses eight synonyms and MW3 nine, but they agree on only three: *inquire, interrogate, question*. Moreover, there is little agreement between the synonyms in MW3 and RHD2 and those listed in the index of *Roget's International Thesaurus, Fourth Edition* (Chapman, 1977).

==== INSERT TABLE 1 ABOUT HERE ==

From these lists, it seems that the editors do not have a set of generally accepted principles for deciding when two words are synonymous. It may seem intuitively obvious that *inquire* is a synonym of *ask*, and *request* a synonym of *ask for*, but our intuitions are probably less sure about the status of *appeal, catechize, charge*, and *require* as synonyms of *ask*.[1]

1. Given how much disagreement there is among these three sources (RHD2, MW3, and Roget's), it is quite striking just how little disagreement there is between *Chambers 20th Century Thesaurus* (Seaton et al., 1986) and *The New Collins Thesaurus* (McLeod, 1984). The Collins Thesaurus lists 25 synonyms for *ask*, all of which can be found in the Chambers Thesaurus. If these sources are truly independent, then we would have to conclude that the notion of synonymy found in an alphabetic thesaurus is more stable than that found in an American unabridged dictionary and in a Roget's thesaurus. It seems more likely, though, that one or both of the books is perpetuating some other lexicographer's intuitions, or perhaps an accretion of lexicographic intuitions, which may or may not reflect true facts about the language.

3.2 Distinguishing Near Synonyms in Reference Works

In addition to identifying synonymous pairs such as *request* and *ask for*, many dictionaries also attempt to describe the differences. Even the most 'synonymous' words are not identical in meaning; you cannot arbitrarily substitute one for the other and expect to preserve the meaning. Although many dictionaries attempt to say why substitution fails, they are not always successful. For example, given the following definition from RHD2, one might incorrectly conclude that *request leniency* and *ask for leniency* are about equally salient. Worse, one might conclude that *asking for it* has approximately the same meaning as *requesting it*. Miller and Gildea (1987) have observed that schoolchildren often misuse dictionaries in just this way, and inappropriately substitute nearly synonymous words into example sentences.

> **ask 10.** to request or petition (usually fol. by *for*): *to ask for leniency; to ask for food*. **11. ask for it**, to persist in an action despite the probability of an unfavorable result: *You couldn't feel sorry for him, because he was asking for it when he continually provoked her.* (RHD2)

These definitions are not unusual or exceptionally misleading. Equally commonplace is the following pair of definitions from *Webster's Ninth New Collegiate Dictionary* (henceforth MW9), which refer to each other (in a somewhat circular fashion), and provide little indication of what either word means, let alone how they differ:[2]

> **ask** *vi* **... 2 :** to make a request <asked for food> (MW9)

> **request 1 :** to make a request to or of **2 :** to ask as a favor or privilege **3 :** *obs* to ask to come or go to something or someplace **4** to ask for (MW9)

Some dictionaries propose that the main difference between *ask for* and *request* is a matter of register. But there can be problems with even a salient distinction such as this. For example, the discussion of the register difference in the following usage note in MW9 suggests that *request* anticipates an affirmative

2. Circularity is generally assumed by logically minded persons to be vicious, although for practical everyday purposes it may well be more adequate than an artificially widened circle (or a refusal to define a set of supposedly 'primitive' terms).

response.

> **ask** REQUEST implies more formality, greater display of courtesy, and anticipation of affirmative response <*request* the cooperation of neighboring towns in the control of Dutch elm disease> <*request* a meeting to discuss common problems and the possibility of mutual help>...

By contrast, our data, discussed in section 5 below, suggests that, in some circumstances at least, people are more likely to be polite and use 'request' if there is a significant chance of a *negative* response. When there is an ''anticipation of an affirmative response,'' one need not be diplomatic, so the neutral verb *ask for* is appropriate, or even a more forceful verb such as *require* or *demand*.[3]

The ''anticipation of affirmative response'' phrase may have been a slip of the pen, or it may be based on boundary cases rather than central and typical examples of usage. If the latter is the case, it is indicative of a more serious problem. We cannot be sure that synonym studies in dictionaries are providing good guidance on appropriate conditions of use. The authenticity of the examples may be beyond doubt, but their naturalness (or typicality) is another question. This paper discusses a number of statistical tools which are intended to address this problem among others. In particular, the *t*-test can be used to find words that are more typically used with one synonym than another. Thus, one difference between *request* and *ask for* is that *anonymity* is more likely to be *requested*, and *leniency* is more likely to be *asked for*, at least on the evidence of usage in the Associated Press newswire. Similarly, *release, freedom*, and *withdrawal* are more likely to be *demanded* than *requested* or *asked for*. These patterns might help justify the conclusion that *demand* is a very strong synonym of *ask*, while *request* is rather a weak one. By studying the statistical distribution of terms in this way, we hope to be able to help sharpen up the description of subtle distinctions among close synonyms.

---

3. The MW9 synonym study for *ask* does not mention the synonym *demand*. *Demand* is accorded a separate synonym study, in which it is contrasted with *require, claim*, and *exact*, but not *ask, ask for*, and *request*. Since there is pretty clear agreement among dictionaries that *request* contains an element of politeness and *demand* contains an element of peremptoriness, it is surprising that they were not studied together as part of a contrasting set. A more systematic, factually based set of procedures for choosing which words to contrast with one another could be a boon to future lexicographers.

The *sub*-test proposed here takes a first step in the direction of deriving sets such *ask for*, *request*, and *demand* by statistical means from a corpus of raw texts. It builds upon our previous work, in which we have used other statistics such as mutual information and *t*-scores to compare and contrast the distribution of words in context.

4. Mutual Information: A Measure of Association

4.1 Using Mutual Information to Summarize a Concordance

The mutual information statistic can be used to identify some interesting objects associated with the two verbs *request* and *ask for*. We discussed the use of the mutual information statistic in (Church and Hanks, 1990) as a tool for identifying interesting associations among words in a corpus. Mutual information is an important statistic in information theory, and can be found in the opening chapter of most textbooks on the subject, including (Fano, 1961, p. 28). The statistic has been very useful in a wide range of applications over the past 40 years, and continues to be important. Some applications in information retrieval, for example, are discussed in van Rijsbergen (1977), while applications in speech recognition are discussed in Jelinek (1985).

Suppose that we saw the sequence ''*requested and*'' showing up a number of times in the concordances to *requested* and wanted to know if there might be a linguistically interesting pattern. Some sequences in the concordances are interesting (e.g., ''*requested anonymity*''), but others such as ''*requested and*'' are not, even though they may be quite frequent. Mutual information can help distinguish the more interesting sequences from the less interesting ones by comparing the joint probability of the sequences with chance. Pairs of words with high mutual information scores are likely to be interesting to a lexicographer.

Let us go through the ''*requested anonymity*'' and ''*requested and*'' examples. In our corpus of N = 44,344,077 words from the 1988 Associated Press newswire, we observed 161 instances of ''*request anonymity*.'' Thus, the joint probability of ''*requested anonymity*'' is $161/N \approx 3.6$ per million. Mutual information compares this probability with chance: the probability of ''*requested*'' times the probability of ''*anonymity*.'' Since we have 1419 instances of ''*requested*'' and 4764 instances of ''*anonymity*,'' chance is $1419/N \times 4764/N \approx 0.0034$ per million. If we now compare the joint (3.6 per million) to chance (0.0034 per million), we see that the joint is much larger than chance ($3.6/0.0034 \approx 1059$), indicating that

''*requested anonymity*'' is probably very interesting, and it might be worthwhile to ask a lexicographer to see if the apparent pattern is linguistically significant.

Since the ratio of the joint probability $P(x,y)$ to chance $P(x)\ P(y)$ tends to be quite large, mutual information $I(x;y)$ expresses the ratio as a logarithm.

$$I(x;y) \equiv \log_2 \frac{P(x,y)}{P(x)\ P(y)}$$

Thus, $I(\textit{requested}; \textit{anonymity}) \approx \log_2 1059 \approx 10$, which is a relatively large score. In contrast, $I(\textit{requested}; \textit{and}) \approx -0.2$, because the joint ($22/N \approx 0.50$ per million) is slightly less than chance ($1419/N \times 793296 \approx 0.57$ per million). Since the mutual information is small (near zero), it is unlikely that ''*requested and*'' is interesting, and there is no statistical reason why a lexicographer should look at it any further.

==== INSERT TABLE 2 ABOUT HERE ==

In (Church and Hanks, 1990), we argued that a table of mutual information values such as these could be used as an index to a concordance. Mutual information can help us decide what to look for in the concordance; it provides a ''quick'' summary of what company our words keep (Firth, 1957).

==== INSERT TABLE 3 ABOUT HERE ==

4.2 Applying Mutual Information to the Output of a Parser

It is also worthwhile to consider more interesting contexts. Suppose that we wanted to look at the objects of the verb *request*. Then, instead of looking at the word immediately after *request* in a concordance, we should look at the word in the direct object position in a parse tree.

We used the (imperfect) Fidditch parser to parse the 44 million word 1988 corpus (in about 16 days of computer time on a Sun4) and extract approximately 4.1 million verb-object (VO) pairs, as discussed in (Church, Gale, Hanks, Hindle, 1990).[4] The parser attempts to undo the effects of various syntactic

transformations (e.g., wh-movement, passive, raising, equi) and also various morphological processes, especially regular inflection. Although the parser makes many mistakes, we have found that the VO pairs are good enough to support surprisingly powerful inferences. For example, Table 4 shows that *request anonymity* is a lot more interesting than *request it*.[5]

==== INSERT TABLE 4 ABOUT HERE ==

In order to highlight the interesting associations, let us restrict our attention to statistically significant VO pairs (as defined in the next section). By filtering out the less interesting pairs, we are left with much higher grade ore. The table below lists the significant objects for *request*, sorted by mutual information scores (which are given in parentheses). Note that a large percentage of the statistically significant objects are also linguistically significant, though there are a few mistakes such as *Syrian*, most of which result from parsing or tagging errors. The significant objects are relatively high-grade ore, which is much more useful to a lexicographer than a raw concordance of all occurrences of these verbs and their objects.

4. The parser uses *UNKNOWN* when it can't find an object, either because there isn't one (the sentence is intransitive), or because of a parsing error. If we look at enough sentences, the statistics can highlight the interesting patterns despite a certain number of parsing errors. In this case, the large number of instances of *request UNKNOWN* (348) in contrast to the complete absence of *ask for UNKNOWN* (0) is because *request* is often used intransitively (with a that-complement, for example) unlike *ask for*.

5. $I(\textit{request/V};\ \textit{anonymity/O}) \approx \log_2 \frac{175/N}{1654/N \times 380/N} \approx 10.2$, where $N \approx 4.1$ million VO pairs.

request (59 significant objects): anonymity (10.2), Syrian (9.7), recount (9.2), reassignment (9.0), asylum (7.9), extradition (7.7), General (7.6), Maikovskis (7.5), secretary-general (7.3), mediation (7.2), postponement (7.1), consultation (7.1), extension (6.3), retirement (6.2), permission (6.2), assistance (6.2), mistrial (6.1), visa (5.8), intervention (5.6), injunction (5.6), Office (5.6), delay (5.4), transfer (5.2), hike (5.2), dismissal (5.1), meeting (5.1), privilege (5.1), exemption (5.0), hearing (4.8), counsel (4.7), copy (4.6), probe (4.6), help (4.5), study (4.5), protection (4.4), information (4.3), data (4.2), inspection (4.2), appointment (4.2), investigation (4.1), inquiry (4.1), warrant (4.0), equipment (3.9), aid (3.8), document (3.6), review (3.6), sentence (3.6), release (3.5), than (3.4), treatment (3.2), session (3.1), trial (3.0), report (2.5), order (2.4), increase (2.0), action (2.0), $ [i.e., a specified sum of money] (1.4), them (1.2), it (0.8)

This list contains a large number of nouns that denote actions or states of affairs (*anonymity, recount, reassignment, asylum,* etc.) and a small number of nouns that denote human agents (*secretary-general, counsel*) or agencies (*company, Office*). The latter group are generally found with an infinitival complement, as in: ‘‘The Senate has requested the General Accounting Office to investigate terms of the deal...’’ Future dictionaries might want to split these two sets into two senses on syntactic grounds. Thus, there would be one sense for the syntactic pattern *request something* (e.g., *anonymity, recount, reassignment, asylum,* etc.) and a second sense for the syntactic pattern *request somebody* (e.g., *counsel, company*, etc.) *to do something*. Only the first of these two senses corresponds to *ask for*. The second corresponds to *ask somebody to do something*. By the same token, there are many uses of *ask* that do not overlap with *request*. Even highly synonymous words are not substitutable in all of their uses.

It might be interesting to compare the list above with a similar list of objects for *ask for*:

ask for (85 significant objects): glass_water (10.1), reconsideration (9.8), leniency (9.6), rehearing (9.5), forgiveness (8.9), asylum (8.7), recount (8.6), autograph (8.2), pity (7.8), mistrial (7.5), help (7.2), extradition (7.1), mercy (7.1), acquittal (6.9), reimbursement (6.7), clemency (6.6), reply (6.6), appropriation (6.5), delay (6.5), extension (6.2), permission (6.1), injunction (6.1), ransom (6.1), temporary (6.1), dismissal (5.9), clearance (5.7), bail (5.6), pardon (5.6), stay (5.5), assistance (5.5), explanation (5.4), exemption (5.4), proof (5.4), donation (5.4), expression (5.3), restraint (5.2), credential (5.1), advice (5.0), protection (5.0), contribution (5.0), timetable (5.0), sentence (4.9), sample (4.8), review (4.8), direction (4.8), copy (4.8), meeting (4.7), resignation (4.7), probe (4.6), interview (4.3), cooperation (4.3), investigation (4.2), trial (4.2), suggestion (4.1), aid (4.0), support (4.0), hearing (4.0), period (3.8), comment (3.8), release (3.7), increase (3.7), information (3.7), penalty (3.7), approval (3.5), money (3.4), damage (3.3), trouble (3.3), amount (3.3), loan (3.2), 0 [trace] (3.1), vote (3.1), room (3.1), commitment (3.0), anything (3.0), study (2.7), $ [money] (2.7), term (2.7), detail (2.5), test (2.4), talk (2.2), name (2.1), more (1.9), report (1.8), time (1.8), program (1.8)

Note that these two sets have a large overlap, suggesting that *ask for* and *request* have similar distributions, and perhaps similar meaning as well. We will return to this point shortly, when we discuss substitutability.

ask for & request (28 significant objects): $ [money], aid, assistance, asylum, copy, delay, dismissal, exemption, extension, extradition, hearing, help, increase, information, injunction, investigation, meeting, mistrial, permission, probe, protection, recount, release, report, review, sentence, study, trial

5. *t*-score: a Measure of Contrast

As we said before, the similarity between *ask for* and *request* is clearly set out in most dictionaries, but the dictionaries are less successful in capturing the differences between *ask for* and *request*. In previous work (Church, Gale, Hanks, Hindle, 1990), we discussed the use of the *t*-test for finding subtle differences between near synonyms. In that case, we looked at the difference between *strong* and *powerful* and concluded that, among other things, *strength* is intrinsic whereas *powerfulness* is extrinsic. Thus, for example, any worthwhile politician or cause can expect *strong supporters*, who are enthusiastic, convinced, vociferous, etc. But far more valuable are *powerful supporters*, who will brings others with them. They are also, according to the AP news, much rarer—or at any rate, much less often mentioned.

A VO pair is considered significant if and only if $t > 1.65$, where

$$t \equiv \frac{P(V,O) - P(V)P(O)}{\sqrt{\sigma^2(P(V,O)) + \sigma^2(Pr(V)Pr(O))}}$$

In this calculation, the probabilities and their variances (σ^2) are computed using the ELE (Expected Likelihood Estimator) method, as described in (Church, Gale, Hanks, Hindle, 1990). That is,

$$P(V,O) \approx \frac{freq(V,O) + 0.5}{N + V/2}$$

and

$$\sigma^2 P(V,O) \approx N\ P(V,O)$$

where N is the size of the corpus (4.1 million), and V is the number of different VO pairs (600,000). In most cases, the ELE produces probability estimates that are very similar to the MLE (maximum likelihood estimate) which omits the 0.5 in the numerator and the $V/2$ in the denominator. We tend to use the ELE when computing significance (e.g., *t*-scores), and MLE when computing maximum likelihood (e.g., mutual information).

The *t*-scores in Tables 5-7 highlight the differences among *ask for*, *request*, and *demand*. There are many interesting points here.[6] In particular, *anonymity* shows up with a significant *t*-score in all three tables. Table 5 shows that *anonymity* is more likely to be *requested* than *asked for* (t = 13.08 standard deviations), and Table 6 shows that *anonymity* is more likely to be *demanded* than *asked for* (t = 12.35 standard deviations). This much is fairly obvious, and can be seen fairly readily in the concordances. One doesn't need fancy statistics to highlight the obvious. However, Table 7 indicates that *anonymity* is more likely to be *requested* than *demanded*. This generalization would be very hard to spot in the concordances, since both are very common. In fact, they are about equally common (175 instances of *request anonymity* vs. 165 instances of *demand anonymity*). The difference is significant, however, because *request* is much less

6. Two items (*which* and *than*) were removed from the tables. These were introduced by errors in the parser, which we expect will be fixed in the near future.

common than *demand* (1654 vs. 4154), and therefore, by chance, one would not have expected so many instances of *request anonymity*.

A look at the concordances (selection below) suggests that *anonymity* is often important to journalistic sources—so important, in fact, that the neutral verb *ask for* is apparently not felt to be appropriate. Comparison with other typical objects (Table 7) suggests that sources may *demand* anonymity when they feel they have a right to do so, but *request* it when they are less sure that it will be granted. Sources that *demand anonymity* are more likely to anticipate an ‘‘affirmative response’’ than those that merely *request* it. This is the basis on which we question the suggestion in MW9 (cited above) that *request* implies an ‘‘anticipation of an affirmative response’’.

Concordances from 1989 Associated Press

One U.S. official , who	requested anonymity	, said recently that it would be difficu
right now , ’’ said the official , who	requested anonymity	, in a telephone interview . *E* *S* ‘‘
A Belgian official , who	requested anonymity	, said there were deaths but declined to
basis , ’’ said the spokesman , who	requested anonymity	. *E* *S* The job losses will be spread
in Vienna , said the NATO source , who	demanded anonymity	. *E* *S* He refused to reveal the number
by TASM , ’’ said an official , who	demanded anonymity	. *E* *S* West Germany is thought to be r
, ’’ said the senior official , who	demanded anonymity	. *E* *S* ‘‘ On the other hand , his diff
flag . *E* *S* A fishery official , who	demanded anonymity	, said North Korea had no agreement to ca

==== INSERT TABLE 5 ABOUT HERE ==

==== INSERT TABLE 6 ABOUT HERE ==

==== INSERT TABLE 7 ABOUT HERE ==

We have often been asked to explain when it is better to use mutual information and when it is better to use *t*-scores. Our view is that one cannot expect a self-organizing silver bullet that will do all of a lexicographer’s work without human intervention. Different statistical tests provide different insights. Often the best strategy is to try a number of tests and then decide what is more helpful and what is less helpful, because the various tests tend to have different strengths and weaknesses. In particular, *t*-scores are

more likely to find high frequency pairs like ‘‘eat food’’ and ‘‘drink water,’’ which are blindingly obvious and more likely to be useful to a computer than a human being, whereas mutual information is more likely to find less frequent collocations such as ‘‘hoard food’’ and ‘‘boil water,’’ which are more specific to the words involved.

The two tests have very different failure modes. *t*-scores have a failure mode of tending to show function words and other very high frequency words that may be too compositional to be of interest, whereas mutual information scores have a failure mode of tending to show low frequency pairs that are too specific to a particular corpus and may not generalize very well. It is often a good idea to intersect the two measures and look at pairs that are high by both measures. This conservative procedure tends to guard against both types of errors, though of course it tends to miss many interesting pairs.

6. Substitutability

Ideally, we would like to be able to measure synonymy directly from corpus data. Unfortunately, since we don’t know how to do this, we have to settle on an approximation which we call *substitutability*. The basic notion of substitutability is to ask: if we substitute the word *x* for the word *y* in some context, what is the chance that we shall end up with a reasonably coherent sentence of English? For instance, if we randomly select concordance lines containing *ask for* and replace the phrase *ask for* with *request*, do we end up with a sensible concordance line more often than we might expect by chance?

By posing the question in this way, we are replacing questions about synonymy (a semantic notion) by questions about textual substitutability (a distributional notion). Whereas two broadly synonymous items may be substituted for one another without affecting either the propositional content of the discourse or its lexicosyntactic structure, two items are substitutable (in our terms) merely in terms of their relationships with other items in the lexicosyntactic structures in which they occur. The proposed *sub*-test can be used to identify sets of items which are potentially synonymous or co-hyponymous, or even antonymous. A lexicographer needs then to use judgement, discretion, intuition, and analysis of the output in order to determine the precise lexical relationship obtaining between the pairs. The *sub*-test may also be used to reject cases of purported synonymy by simply and clearly demonstrating a lack of lexicosyntactic fit or substitutability. Substitutability may then be defined as the phenomenon of two items appearing in

discourse to occupy the same lexicosyntactic space.

For pragmatic reasons, we will take a simpler view of context. Instead of using a whole concordance line, we will consider a significant VO pair. Now, the question is: if we randomly select a significant VO pair with the verb *ask for* and replace the verb with *request*, do we find that we end up with a sensible pair (e.g., another significant VO pair) more often than we would expect?

Section 3.2 listed 85 significant objects for *ask for*, 59 significant objects for *request*, and 28 in the overlap. The large overlap is fairly compelling evidence for concluding that *ask for* and *request* have similar distributions. One might suspect further that the two verbs would also have similar meanings, though, of course, distributional evidence alone cannot be used conclusively to prove semantic regularities, as Wilks (1990: 87) observes.

How many should we have expected by chance? To answer this question, we need a model of chance. We have decided to consider what would happen if we substituted *request* in a randomly selected VO pair. If we had a pair *xz* and we replaced *x* with *request*, with what probability would we get another significant pair? We start by counting the number of significant pairs *xz* such that *request z* is also a significant pair. We find that there are 2382 such significant VO pairs, out of a total of 75,151 significant VO pairs. Thus, given a randomly selected pair, there is a probability of 2382/75,115 that the verb can be replaced with *request*. Since we have 85 *ask for z* pairs, we expect that $85 \times 2382/75,115 \approx 2.69$ of these objects will also appear after *request*, under the null hypothesis that there is no interesting linguistic relationship between *ask for* and *request*.

This expectation of 2.69 should be compared with the observation of 28. Since we observed 28 objects in the overlap, which is more than ten times the expectation of 2.69, it is very likely that the null hypothesis is wrong, and that there is in fact an important linguistic relationship between *request* and *ask for*. We can make this argument precise by formulating the difference between 28 and 2.69 as a *t*-score.

$$t \approx \frac{28 - 2.69}{\sqrt{28\ (1 - 28/85) + 2.69\ (1 - 2382/75,115)}} \approx 5.47$$

==== INSERT TABLE 8 ABOUT HERE ==

In this case, we can very confidently reject the null hypothesis and assume that there almost certainly must be an interesting linguistic relationship to explain why there is such an overlap in the distribution between *request* and *ask for*. In general, we need a *t*-score of 1.65 standard deviations or more in order to reject the null hypothesis with 95% confidence.

Some verbs that substitute for *ask for* and *request* are shown below with *t*-scores in parentheses.

> Some verbs that substitute for *ask for*: request (6.3), seek (5.7), grant (4.1), obtain (3.4), demand (3.1), need (3.0), receive (3.0), withhold (3.0), secure (2.7), await (3.7), deserve (2.6), extend (2.4), pledge (2.4), offer (2.4), agree to (2.2), provide (2.2), collect (2.2), get (2.1), lend (2.1), submit (2.1), delay (2.0)

> Some verbs that substitute for *request*: ask for (5.5), seek (4.3), grant (3.3), await (2.7), obtain (2.5), need (2.5), approve (2.4), demand (2.4), receive (2.3), require (2.2), conclude (2.2)

The verb *seek* shows up in both lists above. It is interesting because it shares a very similar distribution with both verbs (*ask for* and *request*), and yet intuitively *seek* does not seem to be a very good synonym for either of the other two verbs. (It should be noted, though, that *seek* is listed under *ask* in several thesauruses including Roget's, Collins, and Chambers.) It seems that *ask for* and *request* may be more synonymous with each other than either is with *seek*.

Consider some of the verbs that substitute for *seek*:

> ask for (7.9), demand (7.0), grant (6.4), request (6.4), obtain (6.1), receive (5.3), secure (5.2), negotiate (4.8), offer (4.6), guarantee (4.2), provide (4.1), need (4.1), result in (4.0), gain (3.9), withhold (3.9), accept (3.7), win (3.7), agree to (3.7), approve (3.7), favor (3.6), get (3.6), achieve (3.6), oppose (3.5), lead to (3.4), announce (3.4), give (3.2), propose (3.2), recommend (3.1), await (3.1), ensure (3.0)

Note that the same verbs keep coming up. For example, *demand*, *grant*, and *obtain* are found to substitute for *seek*, *request*, and *ask for*. This pattern is probably not an accident. We might suggest that there is a semantic class of *sought-after* objects, and that these *seeking* verbs tend to take members of this class as direct objects.

7. Substitutability and Hyponymy

Note that this definition of substitutability is not symmetric. It is more likely that *request* can be replaced with *ask for* than the reverse, because 28/59 is larger than 28/85. In general, it is easier to substitute a less frequent word for a more frequent word; the numerator (the overlap) will be the same in either order, but the denominator will reflect the different frequency of the two words. The same trends are also found in the *t*-scores. Note, for example, that the word with more significant objects tends to have a smaller *t*-score.

==== INSERT TABLE 9 ABOUT HERE ==

It is especially common to find that a less frequent word (such as *pledge*) can be replaced with a more frequent word (such as *ask for*), but not the other way around. The reason is that 6 (the overlap between the *pledge*-able objects and the *ask for*-able objects) is a much larger percentage of 19 (the number of *pledge*-able objects) than of 85 (the number of *ask for*-able objects). Table 10 lists a number of pairs where *x* can be substituted for *y*, but not the other way around. The columns labeled e1 and e2 are the expectations of substituting x for y, and y for x, respectively. The columns labeled t1 and t2 are the corresponding *t*-scores for the two substitutions. Note that t1 is significant and t2 is not, in each case.

==== INSERT TABLE 10 ABOUT HERE ==

It is not always clear what to make of these pairs. There is often a semantic relationship of some kind, but not always one that fits nicely into a familiar category such as synonymy, antonymy, and hyponymy. For example, it is hard to explain why *pledge* and *contribute* are substitutable. Perhaps, the relationship is mediated through a third word such as *give*. That is, one might claim that *pledge* is related to *give* by hyponymy, which, in turn, is related to *ask for* by antonymy.

In many such cases, however, the relationship may be better explained as one that is temporal, sequential, or narrative. For example, before *asking for* something—say, asylum—people must feel that they *need* it. Then they *ask for, request, seek,* or *demand* it (depending on circumstances, temperament, attitude, etc.). Later, someone else *offers, grants,* or *agrees to* it, or alternatively *withholds* it. As a result, the original applicants *get, obtain,* or *receive* it (or not, as the case may be). In the interim, they may be said to be *awaiting* it, and even to *deserve* it. Note that all these verbs are highly substitutable for *ask for* in our AP corpus.

This ‘temporal’ paradigmatic relationship is more often discussed in textbooks on rhetoric than those on semantics. It appears that the temporal paradigmatic relationship seems to be far more prominent among verbs than objects, even when the lexical semantics of the words in question would lead us to expect otherwise. For example, consider the verbs *nominate* and *impeach* (and *assassinate*!) which are all highly substitutable for *elect*: the relationship is clearly temporal. However, the corresponding objects, *nomination* and *impeachment* (and *assassination*), are not highly substitutable for *election*, indicating that the temporal relation does not work as well for the objects as it does for the verbs.

Some verbs that substitute for *elect* (with a *t*-score and typical object in parentheses): accuse (4.6 *him*), appoint (4.6 *him*), choose (4.3 *successor*), name (4.3 *him*), nominate (3.8 *him*), tell (3.8 *him*), become (3.6 *president*), urge (3.4 *president*), ask (3.4 *him*), select (3.3 *delegate*), convict (3.3 *him*), pick (3.0 *successor*), say (3.0 *Tuesday*), advise (3.0 *him*), defeat (2.9 *Bush*), arrest (2.8 *him*), invite (2.8 *him*), quote (2.8 *him*), assassinate (2.7 *him*), beat (2.7 *him*), embarrass (2.5 *him*), inform (2.5 *him*), hire (2.5 *him*), oust (2.5 *Manigat*), impeach (2.5 *Mecham*), run_against (2.4 *Mitterrand*), brief (2.3 *president*), convince (2.2 *I*), treat (2.2 *him*), criticize (2.2 *him*), swear (2.2 *successor*), greet (2.2 *him*), fire (2.2 *him*), shoot (2.2 *him*), marry (2.2 *who*), address (2.2 *delegate*), kill (2.2 *him*), join (2.1 *him*), expel (2.0 *him*), indict (2.0 *him*), detain (2.0 *him*), endorse (2.0 *Bush*), allow (2.0 *him*), inaugurate (2.0 *Manigat*), portray (2.0 *Dukakis*), bear (2.0 *him*), run_for (2.0 *president*), talk_to (2.0 *president*)

Some objects that substitute for *election* (with a *t*-score and typical verb in parentheses): (3.2 *hold*) meeting, (2.8 *win*) debate, (2.7 *call*) action, (2.7 *oversee*) implementation, (2.6 *hold*) ceremony, (2.6 *hold*) referendum, (2.5 *win*) vote, (2.5 *schedule*) execution, (2.5 *hold*) hearing, (2.4 *win*) primary, (2.4 *face*) deportation, (2.3 *supervise*) withdrawal, (2.3 *win*) release, (2.2 *ensure*) return, (2.2 *schedule*) arraignment, (2.1 *oversee*) project, (2.1 *disrupt*) voting, (2.1 *hold*) session, (2.1 *call*) strike, (2.1 *conduct*) test, (2.1 *hold*) talk, (2.0 *hold*) trial, (2.0 *hold*) demonstration, (2.0 *disrupt*) event, (2.0 *win*) contest, (2.0 *seek*) dismissal, (2.0 *schedule*) launch

There are some interesting differences between substitutable verbs and substitutable objects. As we have just seen, it appears that verbs are more likely to exhibit a temporal relation. In contrast, it appears that certain physical objects are more likely to exhibit a spatial relationship. For example, the only substitutable

picked out by the technique for *window* is *door*. The substitutables for *field* include *street, store, house, car, school, hospital*, and *building*: what these words have in common is that they denote physical locations.

8. Conclusion

The *sub*-test presented in this paper is one of a set of statistical tools being developed at Bell Laboratories which are intended to make possible more accurate descriptions of the patterns to be found in the tremendous flow of everyday language use. At their heart are the twin concepts of mutual information and word association described in (Church and Hanks, 1990). The mutual information test enables lexicographers to find words in a corpus that collocate with each other significantly more often than they would by chance distribution. The *t*-test is complementary to this, in that it highlights contrasts in the collocates of selected pairs of words: it is particularly useful in the study of near synonyms. However, it does not give any help in picking out groups of words that can be regarded as near synonyms. This is where the *sub*-test comes in.

The *sub*-test sorts words, on the basis of their collocates, into paradigmatic sets. It is highly probable that any word in a particular set can be substituted in context with any other word in the same set and yield a sentence that is meaningful, natural, and plausible. The test assigns a numerical value to this probability. The *sub*-test says nothing about the nature of the semantic relations between members of the set that it picks out. We have observed five types of semantic relations within these sets: hyponymy (*requesting* and *demanding* are ways of *asking for* something), synonymy (*get* and *obtain* have nearly the same meaning), antonymy (*withhold* means the opposite of *grant*), temporal (after something has been *requested* it is normally either *granted* or *withheld*), and spacial (a *window* bears a spatial relation to a *door*).

The potential applications of the *sub*-test tool are many and various. As we study the output, we become aware of the enormously complex network that relates our words to one another. As we saw in the case of *anonymity*, even just three words can present an intriguing web of relationships. Great care and skill will be needed in interpreting the salient features of the sets that are identified by the tool. In lexicography, it can be used to select groups of words for consideration by the compiler of an alphabetic thesaurus, and it may help the compiler to arrange synonyms in order of semantic closeness to the headword in the entry. The tool can also be used to generate lists of pairs for input to *t*-tests, which can in turn be used as source

material by writers of synonym studies (as opposed to lists), of the kind found in MW3 and RHD2. This procedure (*sub*-test followed by *t*-test) can also be useful to lexicographers writing definitions contrastively. There are also potential applications outside lexicography in information retrieval and computational linguistics.

At present, the technique has only been applied to nouns (in direct object position) and verbs. It would be comparatively straightforward to apply it to attributive adjectives in relation to their head nouns, although predicative adjectives present a rather greater challenge.

When applied to the 4.1 million VO pairs derived from the 1988 AP corpus, the technique yields a substantial reference work of some 132 pages of verb entries and 125 pages of noun entries. If a work of this kind were to be prepared from a corpus for practical use, an editor might want to divide up the items in each entry according to the type of relation (synonym, antonym, hyponym, temporal, spatial).

The usefulness of the tool is of course constrained by the input: in particular, the composition (size and balance) of the corpus and the accuracy and delicacy of the parser. Size is a more important constraint than balance. Attempts to run the *sub*-test on small, balanced corpora of only a million words or so yield quite impoverished results. The results given here are derived entirely from 1988 Associated Press Corpus (unless stated otherwise). It will be interesting to compare the results obtained by us with results obtained from other corpora, and to ask the question, ‘‘Which of these facts is true of English, and which only of American journalese or of 1988 news stories?’’ Intuitively, quite a number of the facts discovered by the *sub*-test seem plausible candidates as facts of English. We believe that many of the entries in the ‘automatic thesaurus’ generated by this test may well remain fairly stable as data from other corpora accrue, but this hypothesis has yet to be tested.

The parser used for the work described here was extremely crude. It found only SVO triples, and failed to identify some of them correctly. There was no input from more complex structures. For example (in the case of verb complementation) ditransitives, clausal objects, infinitives and verbal nouns, adverbs, and prepositional arguments were all ignored. No doubt, a fuller account of lexical substitutability could be derived from more sophisticated syntactic input. Nevertheless, interesting results were obtained. It seems likely that even better results will be obtained from more balanced corpora and more sophisticated parsers.

References

Box, G., and G. Tiao (1973). *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, Massachusetts, U.S.A.

Brazil, D. (1985). 'The communicative value of intonation in English' English Language Research, University of Birmingham, England.

Chapman, R. (1977). *Roget's International Thesaurus*, fourth edition, Harper & Row, New York, U.S.A.

Chodorow, M., R. Byrd, and G. Heidorn (1985). 'Extracting semantic hierarchies from a large on-line dictionary,' *Association for Computational Linguistics Proceedings*.

Church, K., and P. Hanks (1990). 'Word Association Norms, Mutual Information, and Lexicography,' *Computational Linguistics*, 16:1.

Church, K., W. Gale, P. Hanks, and D. Hindle (1989). 'Parsing, Word Associations and Typical Predicate-Argument Relations,' *International Workshop on Parsing Technologies*, Carnegie Mellon University.

Church, K., W. Gale, P. Hanks, and D. Hindle (1990). 'Using Statistics in Lexical Analysis,' in U. Zernik (ed.), *Lexical Acquisition: Using on-line Resources to Build a Lexicon,* Lawrence Erlbaum Associates, Englewood Cliffs, New Jersey.

Cowie, A., et al. (eds., 1989). *Oxford Advanced Learner's Dictionary*, Fourth Edition, Oxford University Press, England.

Cruse, D. A. (1986). *Lexical Semantics*, Cambridge University Press, England.

Fano, R. (1961). *Transmission of Information*, MIT Press, Cambridge, Massachusetts.

Firth, J. R. (1957). 'A Synopsis of Linguistic Theory 1930-1955' in *Studies in Linguistic Analysis*,

Philological Society, Oxford; reprinted in F. R. Palmer (ed., 1968). *Selected Papers of J. R. Firth*, Longman, Harlow, England.

Flexner, S., L. Hauck, et al. (eds., 1987). *The Random House Dictionary of the English Language, Second Edition, Unabridged*, Random House, New York, U.S.A.

Gale, W. and K. Church (1989). 'What's Wrong with Adding One?' Statistical Research Report #90, AT&T Bell Laboratories, U.S.A.

Good, I. J. (1953). 'The Population Frequencies of Species and the Estimation of Population Parameters,' *Biometrika*, Vol. 40, pp. 237-264.

Gove, P., et al. (eds., 1961). *Merriam Webster's Third New International Dictionary*, G. & C. Merriam Co, Springfield, Massachusetts, U.S.A.

Halliday, M. A. K., and R. Hasan (1976). *Cohesion in English*, Longman, Harlow, England.

Hindle, D. (1983). 'User manual for Fidditch, a deterministic parser,' Naval Research Laboratory Technical Memorandum 7590-142.

Jelinek, F. (1985). 'Self-organized Language Modeling for Speech Recognition,' IBM Report, also available in A. Waibel and K. Lee (eds., 1990) *Readings in Speech Recognition*, Morgan Kaufmann Publishers, San Mateo, California, U.S.A.

Justeson, J. and Katz, S. (1991). 'Co-occurrences of Antonymous Adjectives and Their Contexts,' *Computational Linguistics*, pp. 1-20.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things*, University of Chicago Press, U.S.A.

Leibniz, G.W. von (1704). 'Table de définitions' in L. Couturat (ed., 1903) *Opuscules et fragments inédits*

de Leibniz, Paris, France.

Lyons, J. (1977). *Semantics*, Cambridge University Press, England.

Miller, G., and P. Gildea (1987). 'How children learn words,' *Scientific American*, vol. 257, No. 3, pp. 94-99.

Mish, F. et al. (eds., 1983). *Merriam Webster's Ninth Collegiate Dictionary*, G. & C. Merriam Co, Springfield, Massachusetts, U.S.A.

Quine, W. V. O. (1960) *Word and Object*, MIT Press, Cambridge, Massachusetts, U.S.A.

Seaton, M., G. Davidson, C. Schwarz, and J. Simpson (1986). *Chambers 20th Century Thesaurus*, W. & R. Chambers, Edinburgh, Scotland.

Ullmann, S. (1962). *Semantics*, Basil Blackwell, Oxford, England.

van Rijsbergen, C. (1977). 'A Theoretical Basis for the Use of Co-occurrence Data in Information Retrieval,' *Journal of Documentation*, Vol. 33, No. 2, pp. 106-119.

Wilks, Y. (1990). 'Combining weak methods in large-scale text processing,' Working Notes, AAAI Spring Symposium Series: Text-Based Intelligent Systems, Stanford, U.S.A.

Table 1

Reference	Synonyms of ‘‘ask’’ and ‘‘ask for’’
RHD2	appeal, beg, beseech, entreat, inquire, interrogate, question, sue
MW3	catechize, examine, inquire, interrogate, query, question, quiz, request, solicit
Roget’s index	charge, demand, encourage, inquire, invite, request, require, seek
RHD2 & MW3	inquire, interrogate, question
RHD2 & Roget’s index	inquire
MW3 & Roget’s index	inquire, request

Table 2

I(x; y)	freq(x, y)	freq(x)	freq(y)	x	y
10.0	161	1419	4764	requested	anonymity
–0.2	22	1419	793296	requested	and

Table 3: A Summary of the Words after *requested*

I(x; y)	freq(x, y)	freq(x)	freq(y)	x	y
10.0	161	1419	4764	requested	anonymity
8.2	14	1419	1529	requested	permission
7.8	5	1419	698	requested	asylum
7.3	5	1419	968	requested	copies
7.1	4	1419	935	requested	detailed
6.8	4	1419	1090	requested	background
6.2	9	1419	3744	requested	documents
6.0	5	1419	2519	requested	protection
5.7	6	1419	3498	requested	additional
5.4	4	1419	2928	requested	meetings
5.0	199	1419	190545	requested	by
5.0	4	1419	4014	requested	emergency
5.0	9	1419	8983	requested	information
4.5	13	1419	17379	requested	political
3.5	5	1419	13538	requested	help

Table 4

I(x; y)	freq(x, y)	freq(x)	freq(y)	x	y
10.2	175	1654	380	request/V	anonymity/O
0.8	49	1654	71474	request/V	it/O

Table 5: What do you typically request and what do you typically ask for?

request				ask for			
t	request	ask for	w	t	request	ask for	w
18.63	348	0	UNKNOWN	–6.32	10	65	help
13.08	175	1	anonymity	–5.10	0	27	0 [trace]
2.49	17	5	study	–5.10	0	27	support
2.47	7	0	secretary-general	–4.24	28	70	$ [money]
2.27	6	0	equipment	–4.01	0	17	comment
2.04	5	0	company	–3.74	7	30	money
1.76	6	1	consultation	–3.02	0	10	contribution
1.76	6	1	data	–3.02	0	10	forgiveness
				–2.85	0	9	leniency
				–2.67	0	8	program

Table 6: What do you typically demand and what do you typically ask for?

demand				ask for			
t	demand	ask for	w	t	demand	ask for	w
12.35	165	1	anonymity	–8.06	0	65	help
9.23	137	7	release	–5.79	5	39	asylum
8.08	77	1	end	–5.17	51	70	$ [money]
7.31	65	1	withdrawal	–5.07	24	48	meeting
7.01	53	0	government	–4.66	5	27	support
5.86	38	0	ouster	–4.35	0	19	delay
5.55	48	2	freedom	–4.33	5	24	sentence
5.42	33	0	arrest	–4.12	0	17	comment
5.32	39	1	pay	–3.98	4	20	trial
4.86	40	2	removal	–3.85	4	19	permission
4.84	27	0	legalization	–3.73	0	14	assistance

Table 7: What do you typically demand and what do you typically request?

demand				request			
t	demand	request	w	t	demand	request	w
9.50	137	6	release	–6.87	165	175	anonymity
9.00	1244	348	UNKNOWN	–6.46	24	65	meeting
8.55	77	0	end	–4.79	0	23	assistance
7.30	65	1	withdrawal	–4.68	31	49	it
6.96	67	2	resignation	–4.58	8	29	information
6.65	48	0	freedom	–4.34	2	21	hearing
6.26	43	0	wage	–4.11	5	22	asylum
6.10	41	0	right	–4.11	1	18	extradition
5.94	53	2	government	–3.98	4	20	permission
5.85	38	0	ouster	–3.98	2	18	report
5.31	39	1	pay	–3.98	1	17	study
4.84	40	2	removal	–3.60	0	13	extension
4.83	27	0	legalization	–3.48	6	18	them
4.73	26	0	autonomy	–3.15	2	12	recount
4.16	33	2	arrest	–3.15	0	10	help

Table 8

ask for *z*	request *z*	observed	expected	*t*-score
85	59	28	2.69	5.47

Table 9: Substitutability is not symmetric

xz	xz & yz	expected	t	x	y
85	28	2.69	5.47	ask_for	request
59	28	2.30	6.25	request	ask_for

Table 10: Some Asymmetric Substitutions

xz	yz	xz & yz	e1	e2	t1	t2	x	y
19	85	6	0.74	2.48	2.40	1.25	pledge	ask_for
41	85	7	1.60	3.27	1.99	1.21	conclude	ask_for
67	85	9	2.61	3.82	1.99	1.52	welcome	ask_for
13	85	4	0.51	2.21	1.94	0.73	contribute	ask_for
56	85	8	2.19	3.74	1.94	1.30	wait_for	ask_for
34	85	6	1.33	1.61	1.87	1.64	postpone	ask_for
35	85	6	1.37	1.90	1.85	1.50	undertake	ask_for
48	85	7	1.87	2.43	1.84	1.54	reveal	ask_for
38	85	6	1.48	3.42	1.77	0.87	promise	ask_for
52	85	7	2.03	3.11	1.76	1.27	recommend	ask_for
28	85	5	1.09	1.53	1.72	1.39	object_to	ask_for

Footnotes

[Note to the editor: the footnotes in the body of the paper are duplicated below.]

1. Given how much disagreement there is among these three sources (RHD2, MW3, and Roget's), it is quite striking just how little disagreement there is between *Chambers 20th Century Thesaurus* (Seaton et al., 1986) and *The New Collins Thesaurus* (McLeod, 1984). The Collins Thesaurus lists 25 synonyms for *ask*, all of which can be found in the Chambers Thesaurus. If these sources are truly independent, then we would have to conclude that the notion of synonymy found in an alphabetic thesaurus is more stable than that found in an American unabridged dictionary and in a Roget's thesaurus. It seems more likely, though, that one or both of the books is perpetuating some other lexicographer's intuitions, or perhaps an accretion of lexicographic intuitions, which may or may not reflect true facts about the language.

2. Circularity is generally assumed by logically minded persons to be vicious, although for practical everyday purposes it may well be more adequate than an artificially widened circle (or a refusal to define a set of supposedly 'primitive' terms).

3. The MW9 synonym study for *ask* does not mention the synonym *demand*. *Demand* is accorded a separate synonym study, in which it is contrasted with *require, claim,* and *exact*, but not *ask, ask for*, and *request*. Since there is pretty clear agreement among dictionaries that *request* contains an element of politeness and *demand* contains an element of peremptoriness, it is surprising that they were not studied together as part of a contrasting set. A more systematic, factually based set of procedures for choosing which words to contrast with one another could be a boon to future lexicographers.

4. The parser uses *UNKNOWN* when it can't find an object, either because there isn't one (the sentence is intransitive), or because of a parsing error. If we look at enough sentences, the statistics can highlight the interesting patterns despite a certain number of parsing errors. In this case, the large number of instances of *request UNKNOWN* (348) in contrast to the complete absence of *ask for UNKNOWN* (0) is because *request* is often used intransitively (with a that-complement, for example) unlike *ask for*.

5. $I(request/V;\ anonymity/O) \approx \log_2 \frac{175/N}{1654/N \times 380/N} \approx 10.2$, where $N \approx 4.1$ million VO pairs.

6. Two items (*which* and *than*) were removed from the tables. These were introduced by errors in the parser, which we expect will be fixed in the near future.