AUTOMATIC SEMANTIC CLASSIFICATION OF VERBS FROM THEIR SYNTACTIC CONTEXTS: AN IMPLEMENTED CLASSIFIER FOR STATIVITY

Michael R. Brent
MIT AI Lab
545 Technology Square
Cambridge, Massachusetts 02139
michael@ai.mit.edu

Abstract

This paper discusses an implemented program that automatically classifies verbs into those that describe only states of the world, such as *to know*, and those that describe events, such as *to look*. It works by exploiting the constraint between the syntactic environments in which a verb can occur and its meaning. The only input is on-line text. This demonstrates an important new technique for the automatic generation of lexical databases.

1 Introduction

Young children and natural language processing programs face a common problem: everyone else knows a lot more about words. Children, it is hypothesized, catch up by observing the linguistic and non-linguistic contexts in which words are used. This paper focuses on the value and accessibility of the linguistic context. It argues that linguistic context by itself can provide useful cues about verb meaning to an artificial learner. This is demonstrated by a program that exploits two particular cues from the linguistic context to classify verbs automatically into those whose sole sense is one describing a state, and those that have a sense describing an event.[1] The approach described here accounts for a certain degree of noise in the input due both to mis-apprehension of input sentences and to their occasional mal-formation. This work shows that the two cues are available and are reliable given the statistical methods applied.

Language users, whether natural or artificial, need detailed semantic and syntactic classifications of words. Ultimately, any artificial language user must be able to add new words to its lexicon, if only to accommodate the many neologisms it will encounter. And our lexicographic needs grow with our understanding of language. A number of current approaches to satisfying the lexical requirements for artificial devices do not involve unsupervised learning from examples. Boguraev and Briscoe (1987) discusses interpreting the information published in on-line dictionaries, while Zernik and Dyer (1987) discuss tutored learning in a controlled environment. But any method that requires explicit human intervention — be it that of lexicographers, knowledge engineers, or "tutors" — will lag behind both the growth of vocabulary and the growth of linguistics, and even with the lag their maintenance will be expensive. By contrast, dictionaries constructed by automated learners from real sentences will not lag behind vocabulary growth; examples of current language use are free and nearly infinite. These observations have led several researchers, including Hindle (1990) and Smadja and McKeown (1990), to begin investigating automatic acquisition of semantics. Hindle and Smadja and McKeown rely purely on the ability of one particular word to statistically predict the occurrence of another in a particular position. In contrast, the approach described here is targeted at particular semantic classes that are revealed by specific linguistic constructions.

2 The Questions

This section discusses work on two linguistic cues that reveal the availability of non-stative senses for verbs. This work attempts to determine the difficulty of using the cues to classify verbs into those describing states and those describing events. To that end, it focuses on two questions:

1. Is it possible to reliably detect the two cues using only a simple syntactic mechanism and minimal syntactic knowledge? How simple can the syntax be? (The less knowledge required to learn using a given technique, the

[1]The input sentences are those compiled in the Lancaster/Oslo/Bergen (LOB) Corpus, a balanced corpus of one million words of British English. The LOB consists primarily of edited prose.

more useful the technique will be.)

2. Assuming minimal syntactic power, how reliable are our two cues in real text, which is subject to performance limitations? Are there learning strategies under which their reliability is adequate?

Section 2.1 describes syntactic constructions studied and demonstrates their relation to the stative semantic class. Sections 2.2 answers questions 1 in the affirmative. Section 2.4 answers question 2 in the affirmative, discusses the statistical method used for noise reduction, and demonstrates the program that learns the state-event distinction.

2.1 Revealing Constructions

The differences between verbs describing states (statives) and those describing events (non-statives) has been studied by linguists at least since Lakoff (1965). (For a more precise semantic characterization of statives see Dowty, 1979.) Classic examples of stative verbs are *know*, *believe*, *desire*, and *love*. A number of syntactic tests have been proposed to distinguish between statives and non-statives (again see Dowty, 1979). For example, stative verbs are anomalous when used in the progressive aspect and when modified by rate adverbs such as *quickly* and *slowly*:

(1) a. * Jon is knowing calculus
    b. * Jon knows calculus quickly

Perception verbs like *see* and *hear* share with statives a strong resistance to the progressive aspect, but not to rate adverbs:

(2) a. * Jon is seeing the car
    b. OK Jon quickly saw the car

Agentive verbs describing attempts to gain perceptions, like *look* and *listen*, do not share either property:

(3) a. OK Jon is looking at a car
    b. OK Jon quickly looked at his watch

The classification program relies primarily on the progressive cue, but uses evidence from the rate adverb cue when it is available.

2.2 Syntactic Requirements for Cue Detection

Consider first how much syntactic analysis is needed to detect the progressive and rate adverb constructions. Initially, suppose that the availability of a non-stative sense is an intrinsic property of a verb not affected by its syntactic context.[2] To detect progressives one need only parse a trivial part of the auxiliary system, which is known to be finite-state. Detecting the rate adverb cue requires determining what the adverb modifies, and that can be trickier. For example, adverbs may appear after the direct object, (4a), and this must not be confused with the case where they appear after the subject of an embedded clause, (4b).

(4) a. Jon fixed the robot quickly
    b. Jon knew his hostess rapidly lost interest in such things

Using simple, finite-state machinery one would be forced to deal with (4b) by recognizing the position of the adverb as ambiguous and rejecting the example. Or one could deploy more sophisticated syntax to try determining the boundaries of embedded sentences. But even the best syntactic parser will fail on truly ambiguous cases like the following:

(5) a. Jon fixed the robot that had spoken slowly
    b. Jon believed the robot that had spoken slowly

The data on rate adverbs were collected using the parsing approach, which required a substantial amount of machinery, but a finite-state approach might do almost as well. (See Brent and Berwick, 1991, for automatic lexical acquisition using simple finite-state parsing.)

2.3 Data on Cues from the Corpus

To test the power of the two proposed cues, the LOB corpus was automatically processed to determine what percentage of each verb's occurrences were in the progressive, and what percentage were modified by rate adverbs. Sampling error was handled by calculating the probability distribution of the true percentage for each verb assuming that the sentences in the corpus were drawn at random from some infinitely large corpus. The overall frequency of the progressive construction was substantially higher than that of the rate adverb construction and so provided more significant data. Figure 1 shows a histogram constructed by summing these distributions of true frequency in the progressive over each of the 38 most common verbs in the corpus.[3] Figure 1 shows that, at least for these most common verbs, there are three and perhaps four distinct populations. In other words, these verbs do not vary continuously in their frequency of occurrence in the progressive, but rather show a marked tendency to cluster around certain values. As will be shown in the next section, the

[2] This is not true in general, as shown by the fact that *think that...* is stative while *think about...* is not.

[3] Histograms that include less frequent verbs have the same general character, but the second local maximum gets somewhat blurred by the many verbs whose true frequency in the progressive is poorly localized due to insufficient sample size.

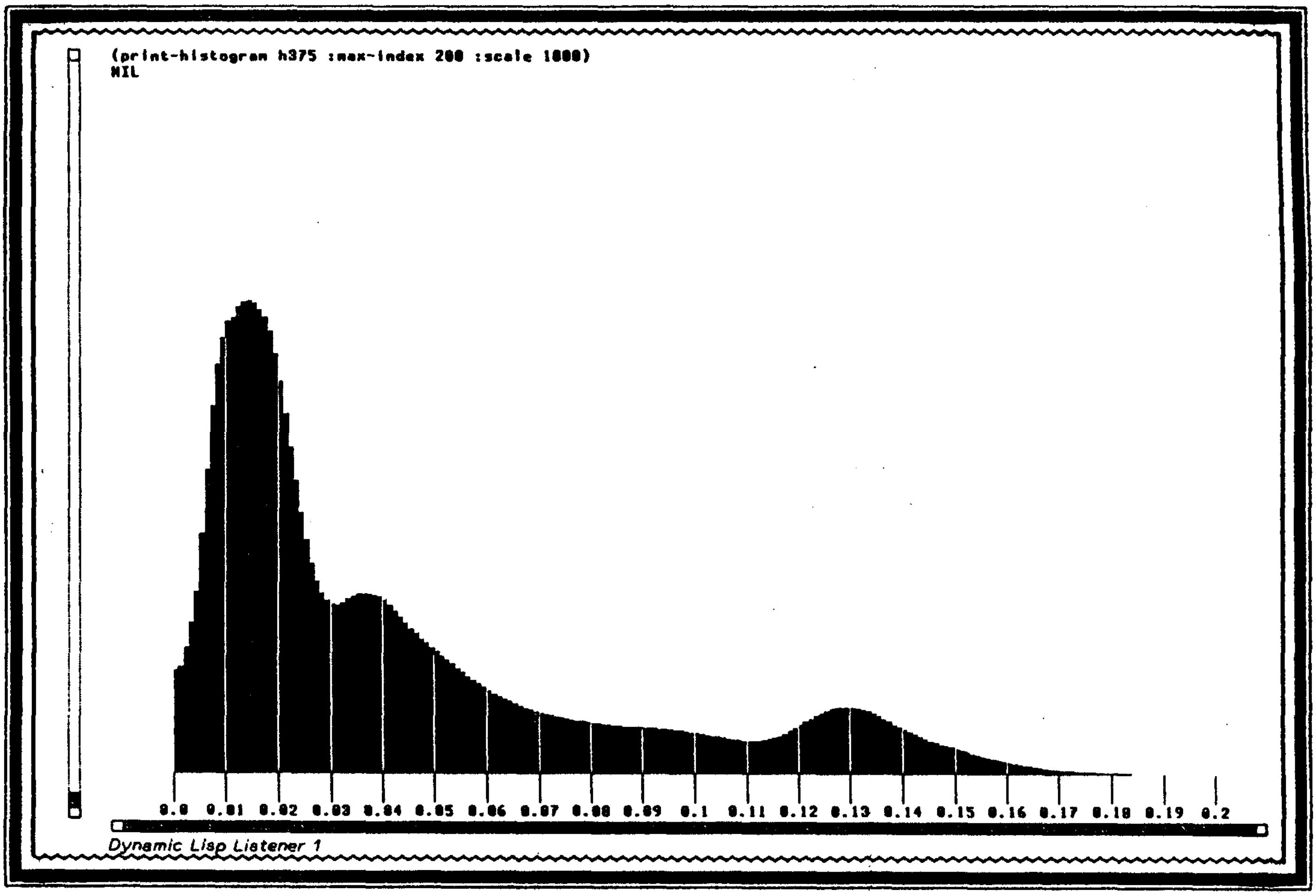

Figure 1: A histogram constructed by summing the probability distributions of true frequency in the progressive over each of the 38 most common verbs in the corpus

stative verbs fall in the first population, to the left of the slight discontinuity at 1.35% of occurrence in the progressive.

2.4 The Classification Program

I implemented a program that attempts to classify verbs into those with event senses, and those whose only meaning describes a state rather than an event. It does this by first detecting occurrences of the progressive and rate adverb constructions in the LOB corpus, and then computing confidence intervals on the true frequency of occurrence of each verb in an arbitrarily large corpus of the same composition. The program classifies the verbs according to bounds, which are for the moment supplied by the researcher, on the confidence intervals. For example, on the run shown in Figure 2, the program classifies verbs which occur at least .1% of the time either in the progressive or modified by a rate adverb, as having an event (non-stative) sense. The classifier acts on .1% bound only if the sample-size is large enough to guarantee the bound with 95% confidence. Accuracy in ascribing non-stative senses according with this technique is excellent — no purely stative verbs are mis-classified as having non-stative senses. In fact, this result is not very sensitive to raising the minimum progressive frequency from .1% to as high as .6% or .7%, since most verbs with non-stative senses yield observed frequencies of at least two or three percent.

Now consider the other side of the problem, classifying verbs as purely stative. Here the program takes verbs that fail the test for having a non-stative sense, and in addition whose true frequency in the progressive falls below a given upper bound with sufficient confidence. The rate-adverb construction is not used, except insofar as the verbs must fail the .1% lower bound, because this construction turns out to be so rare that only a few of the most frequent verbs provide sufficiently tight bounds. The results for identifying pure sta-

```
(classify-stative-non-stative :max-progr-for-stative .0135
                              :min-rate-for-non-stative .001
                              :min-progr-for-non-stative .001)

                    LACKS-NON-STATIVE-SENSE:

KNOW  SEEM  LIKE  BELIEVE  WANT  REMAIN  MEAN  HEAR  REQUIRE  UNDERSTAND  AGREE

                    HAS-NON-STATIVE-SENSE:

WEAR  WAIT  DO  TALK  GROW  GO  STAY  TRY  VISIT  LISTEN  LIE  SIT  PREPARE  FAIL  SEEK  WONDER  FIGHT  WORK  STUDY
BEGIN  DRIVE  WATCH  DEAL  ACT  ENJOY  SETTLE  SMILE  PLAY  DIE  LIVE  RUN  MOVE  STAND  HAPPEN  WALK  CREATE  PROVE
 CAUSE  BREAK  DROP  LOOK  CARRY  FACE  ATTEND  EXPECT  FALL  END  DEVELOP  OFFER  EAT  WRITE  READ  CLOSE  BECOME  P
OINT  DRAW  RETURN  RISE  BUILD  CHANGE  WIN  COME  LEARN  PUBLISH  PICK  ASSOCIATE  GET  PRODUCE  REPLY  PAY  LEAD
INTRODUCE  COMPLETE  REFUSE  SAVE  NOTICE  PULL  AVOID  RECEIVE  SERVE  PRESENT  STOP  OPEN  ENTER  SET  SPEND  WISH
 FORGET  NOTE  ASSUME  PLACE  INCREASE  OCCUR  COMPARE  CONSIDER  SUGGEST  COVER  DISCOVER  SELL  THINK  REGARD  AFFE
CT  KEEP  HOLD  FOLLOW  PUT  MEET  ACCEPT  SEND  HELP  REVEAL  BRING  RAISE  APPEAR  PROVIDE  TAKE  REMEMBER  SPEEK
FINISH  TURN  ASK  REACH  LET  TELL  MAKE  FELL  ANSWER  FIND  LEAVE  ACHIEVE  FEEL  CALL  SHOW  USE  EXPLAIN  GIVE
OBTAIN  DECIDE  SAY  SEE

                    INDETERMINATE:

THROW  REFER  BASE  CHOOSE  CATCH  ARRIVE  MANAGE  SUPPORT  EXIST  BELONG  ARISE  BEAR  NEED  CUT  INTEND  IMAGINE  C
LAIM  FORM  BUY  NE  STATE  KILL  APPLY  REMOVE  REALISE  HOPE  MAINTAIN  JOIN  MENTION  FILL  ADMIT  DEPEND  REPORT
 ALLOW  MARRY  ESTABLISH  INDICATE  LAY  LOOSE  SPEAK  SUPPOSE  REDUCE  REPRESENT  LOVE  INVOLVE  CONTAIN  ADD  START
  BR  INCLUDE  CONCERN  CONTINUE  DESCRIBE
NIL
```

Dynamic Lisp Listener 1

Figure 2: **One run of the stative/non-stative classification program on verbs occurring at least 100 times in the LOB corpus**

tives are also quite good, although more sensitive to the precise bounds than were the results for identifying non-statives. If the upper bound on progressive frequency is set at 1.35%, as in Figure 2, then eleven verbs are identified as purely stative, of the 204 distinct verbs occurring at least 100 times each in the corpus. Two of these, *hear* and *agree*, have relatively rare non-stative senses, meaning to imagine one hears ("I am hearing a ringing in my ears") and to communicate agreement ("Rachel was already agreeing when Jon interrupted her with yet another tirade"). If the upper bound on progressive frequency is tightened to 1.20% then *hear* and *agree* drop into the "indeterminate" category of verbs that pass neither test. So, too, do three pure statives, *mean*, *require*, and *understand*.

It is worth noting the importance of using some sort of noise reduction technique, such as the confidence intervals used here. There are two sources of noise in the linguistic input. First speakers do utter anomalous sentences. For example, the stative verb *mean* occurred one time out of 450 in the progressive. The sentence, "It's a stroke, that was what he was meaning" is clearly anomalous. The second source of noise is failure of the learner to detect the cue accurately. The accuracy of our automatic cue detection detection is described in the following section.

2.5 Accuracy of Cue Detection

Section 2.2 discussed how much structure must be imposed on sentences if the progressive and rate-adverb constructions are to be detected. Section 2.3 showed that the progressive and rate-adverb constructions are indeed reliable cues for the availability of a non-stative sense. This section discusses the accuracy with which these cues can be detected.

It is not practical to check manually every verb occurrence that our program judged to be progressive. Instead, I checked 300 such sentences selected at random from among the most commonly occurring verbs. This check revealed only one sentence that did not truly describe a progressive event. That sentence is shown in (6a).

(6) a. *go*: What that means in this case **is going** back to the war years...

b. *see*: The task was solely **to see how speedily** it could be met...

c. *compare*: ...the purchasing power of the underdeveloped countries in the commonwealth **will rise slowly compared** with that of Europe.

It is not clear how to automatically determine that (6a) does not describe an event of going in progress. Rate adverbs are infrequent enough that it was possible to verify manually all 281 cases the program found. In four of those cases the rate adverb actually modified a verb other than the one that the program chose. Three of these four cases had the structure of (6b), where a wh- relative is not recognized as signaling the beginning of a new clause. This reflects an oversight in the grammar that should be easily correctable. The one remaining case of a mis-attributed rate adverb, (6c), would again require some work, and perhaps substantial syntactic knowledge, to correct. The rate of false positives in cue detection, then can be estimated at about one serious hazard in 300 for both tests.

3 Conclusions

This work demonstrates a promising approach to automatic semantic classification of verbs based only on their immediate linguistic contexts. Some sort of statistical smoothing is essential to avoid being permanently mislead by anomalous and misunderstood utterances, and this work demonstrated the sufficiency of an approach based on binomial confidence-intervals. These methods, in combination with pure collocational methods like those of [Hindle, 1990] and [Smadja and McKeown, 1990], may eventually yield substantial progress toward automatic acquisition of word meaning, or some aspects thereof, by language using devices.

The initial results described here suggest many more experiments, some of which are already underway (see Brent and Berwick, 1991). These include attempting to take into account the ability of local syntactic context to influence a verb's meaning as well as to reveal it. For example, *think that* is stative while *think about* and *think of* are not. Separating these two senses automatically could add substantial power to our classifier. Next, there are many more linguistic cues to verb meaning to be detected and exploited. For example, the ability to take both a direct object and a propositional complement, as in "tell him that he's a fool", reveal verbs of communication. While the progressive cue is not available in Romance languages, the ability to take a direct object and a propositional complement seems to be diagnostic of communication verbs in Romance as well as in English. It would be valuable to demonstrate cues like this on non-English text. It would also be valuable to apply these techniques to a greater variety of input sentences, including transcriptions of mother's speech to their children. Finally, substantially larger corpora should be used in order to enlarge the number of verbs classified. All of these planned extensions serve the goal of automatically classifying thousands of verbs by dozens of different syntactic criteria, and thereby yielding a valuable, adaptable lexicon for natural language processing and artificial intelligence.

References

[Boguraev and Briscoe, 1987] B. Boguraev and T. Briscoe. Large Lexicons for Natural Language Processing: Utilising the Grammar Coding System of LDOCE. *Comp. Ling.*, 13(3), 1987.

[Brent and Berwick, 1991] M. Brent and R. Berwick. Automatic Acquisition of Subcategorization Frames From Free Text Corpora. In *Proceedings of the 4th Darpa Speech and Natural Language Workshop*. Defense Advanced Research Projects Agency, Arlington, VA, USA, 1991.

[Dowty, 1979] D. Dowty. *Word Meaning and Montague Grammar*. Synthese Language Library. D. Reidel, Boston, 1979.

[Hindle, 1990] D. Hindle. Noun classification from predicate argument structures. In *Proceedings of the 28th Annual Meeting of the ACL*, pages 268–275. ACL, 1990.

[Lakoff, 1965] G. Lakoff. *On the Nature of Syntactic Irregularity*. PhD thesis, Indiana University, 1965. Published by Holt, Rinhard, and Winston as *Irregularity in Syntax*, 1970.

[Smadja and McKeown, 1990] F. Smadja and K. McKeown. Automatically extracting and representing collocations for language generation. In *28th Annual Meeting of the Association for Comp. Ling.*, pages 252–259. ACL, 1990.

[Zernik and Dyer, 1987] U. Zernik and M. Dyer. The self-extending phrasal lexicon. *Comp. Ling.*, 13(3), 1987.