Proceedings of The Institute of Acoustics

# LEXICAL STRESS, LEXICAL DISCRIMINABILITY, AND VARIABLE PHONETIC INFORMATION IN SPEECH RECOGNITION

Gerry T.M. Altmann

Centre for Speech Technology Research, University of Edinburgh.

## ABSTRACT

Huttenlocher [2] claims that the phonetic information contained within the stressed syllable of a word is most informative relative to other portions of that word in terms of distinguishing it from rival candidates in the lexicon. Huttenlocher demonstrates this finding within a model of lexical access in which only partial phonetic information is used; specifically, 6 "broad" classes of phonemic unit [3]. This paper describes an attempt to confirm Huttenlocher's findings within a model of lexical access in which *variable* phonetic information is output by a notional acoustic front-end. The model incorporates 44 "FineClass" phonemic units, as well as several superordinate "MidClass" groupings of these phonemes [1]. A number of simulations are described in which different assumptions are made concerning where, within a word, fineclass descriptions might be output by the front-end. Two conditions are compared, in which FineClass descriptions are output at random (with MidClass descriptions elsewhere), and in which they are output only within stressed syllables. The results, which question Huttenlocher's claims, are discussed in relation to the recognition of continuous speech by machine.

## INTRODUCTION

Huttenlocher [2] suggests that certain portions of a word are more informative than others in terms of distinguishing that word from rival candidates in a lexicon. For instance, in a tri-syllabic word like **abandon**, transcribed as **/@ b a n d @ n/**, there are 145 other words in the Oxford Advanced Learner's Dictionary which share the second syllable **/b a n/**, whilst there are 1056 other words which share the third syllable **/d @ n/** (i.e. the syllable **/b a n/** defines an *equivalence class* containing 146 members, whilst the syllable **/d @ n/** defines an equivalence class containing 1057 members). On this basis, we might say that the second syllable of that word is more *informative* than the third syllable. It is also the case that that second syllable is *stressed*. Perhaps, it is a general property of stressed syllables that they are more informative than unstressed sylables. It is this question which Huttenlocher considers.

Huttenlocher bases his own experiments on a model of lexical access in which only partial phonetic information is used; 6 BroadClass phonemes [3]. He calculated the distribution of equivalence classes under essentially three conditions:

(1) Using just BroadClass information alone, for which the transcription of the word abandon becomes **/VOW STOP VOW NAS STOP VOW NAS/**, it turns out that approximately 32% of the lexicon (Merriam Webster's Pocket Dictionary: 20,000 words) is uniquely discriminable.

Huttenlocher then compared the discriminable power of stressed and unstressed

syllables as follows:

(2) For a word like **compare**, transcribed as /STOP VOW NAS STOP VOW/, he altered the transcription so as only to maintain category information in the stressed segment. The transcription thus became /+ STOP VOW/, where a "+" indicates the presence of a syllable, but neither the identity nor the number of its segments. Huttenlocher appended his transcriptions with the (redundant) information that the first syllable was unstressed, and the second syllable stressed.

(3) Finally, he altered the transcription so as only to maintain category information in the unstressed segments: /STOP VOW NAS +/.

On the basis of the distribution of the ensuing equivalence classes, Huttenlocher concludes that the information contained within the stressed syllable is more informative than the information contained elsewhere within the word. The rest of this paper consists of an attempt to replicate this finding within a model of lexical access in which *variable* phonetic information is used: 44 FineClass phonemic categories, and a number of MidClass categories (approximately twice as many as the 6 BroadClasses used by Huttenlocher and Zue).

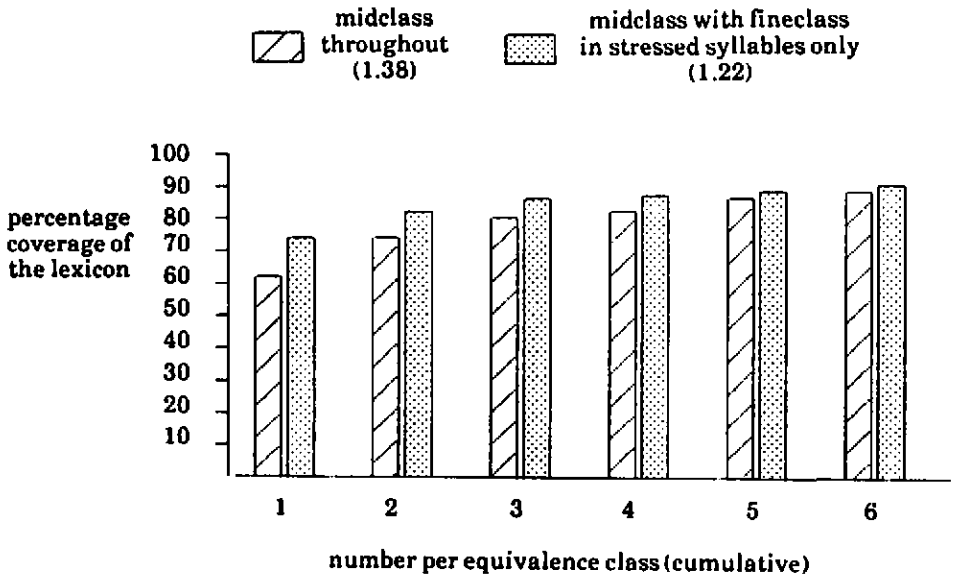
## VARIABLE PHONETIC INFORMATION AND LEXICAL DISCRIMINABILITY

As Huttenlocher and others have pointed out, stressed segments are likely to be acoustically more reliable than unstressed segments. If Huttenlocher's claims hold true, they provide important motivation for the existence of lexical stress: the chances of the hearer recognising a word are maximised if that portion of the word which is most informative is articulated more clearly. In machine recognition terms, the likelihood of the acoustic front-end being able to resolve to the level of FineClasses will be greatest around stressed segments, if it is indeed the case that these are the areas of greatest acoustic reliability. Moreover, given the claim that these segments are most informative relative to other word candidates in the lexicon, it follows that if there is anywhere in the word where we would actively want the front-end to perform well, it would be exactly here.

We might envisage, then, a front-end which was capable of outputing FineClass information but tended to do so only around stressed syllables, and which elsewhere output only MidClass information. A word like abandon might then be transcribed by the system as /CV b a n B CV N/. Preliminary experiments on a 4000-word lexicon suggested that MidClass information alone allows approximately 60% of the lexicon to be uniquely identified. On the basis of Huttenlocher's claims, we could expect a significant increase in discriminability if FineClass information was provided in just the stressed syllables. This was tested using a version of the Oxford Advanced Learner's Dictionary (henceforth "ALD"), containing approximately 18500 words, in which syllable boundaries were marked in the phonological transcriptions (using a maximal onset algorithm). Approximately 33% of the syllables were marked as stressed. Monosyllabic words were marked as unstressed.

Figure 1 shows the distribution of equivalence classes based on transcriptions which are MidClass throughout, and based on transcriptions which are MidClass with FineClass in stressed syllables.

Figure 1. Distribution of equivalence classes for transcriptions which are MidClass throughout, and MidClass with FineClass in stressed syllables only. Figures in parentheses denote mean size of equivalence classes.

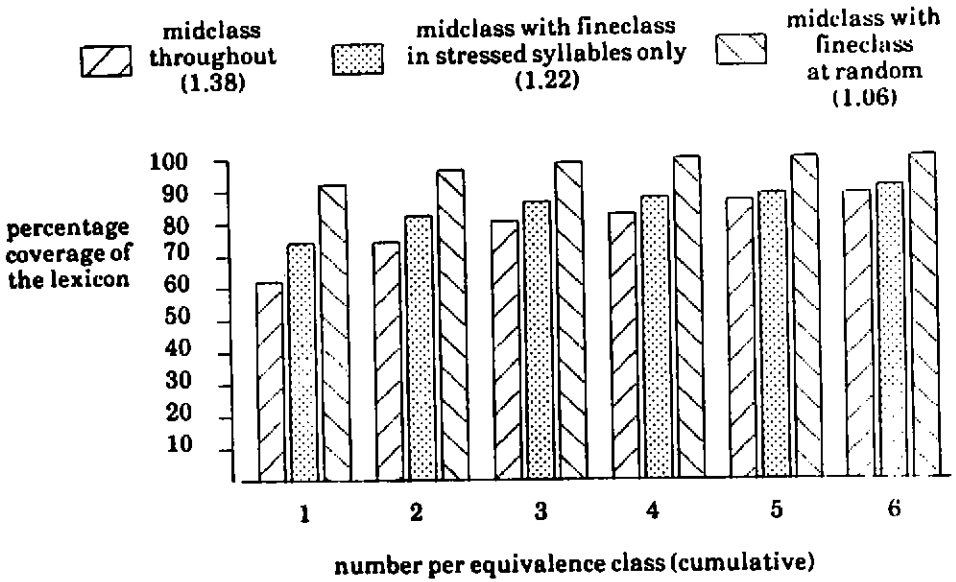



Using MidClass information alone, 62% of the lexicon is uniquely identifiable, and an additional 30% is identifiable if one takes into account equivalence classes which have between 1 and 7 members. The mean equivalence class size was 1.38. In the case of mixed descriptors, with FineClass information in all the stressed syllables, 75% of the lexicon is uniquely identifiable, but equivalence classes with up to 6 members still have to be included to increase this figure to 90% (with the mean class size being 1.22).

All in all, there is an advantage to be gained if the front-end can identify FineClass information in stressed syllables, but the advantage is perhaps not as much as one would expect given the supposed informativeness of these syllables. Furthermore, the fact that 33% of all the segments in the lexicon fell within stressed syllables, and hence were transcribed to the FineClass level, suggests that the payoff is in fact quite small given the amount of extra work required of the front-end.

Perhaps, then, stressed syllables aren't as informative as Huttenlocher would claim. A second simulation was run in which, once again, 33% of the segments in the lexicon were transcribed to the FineClass level and the rest to MidClass, but this time segments were selected *at random*, and not according to whether or not they fell within a stressed syllable. The transcription of the word abandon might thus become /CV B a N d CV n/. If stressed syllables have no special informative status, then we might expect little difference between the two simulations. On the other hand, if stressed syllables are the more informative, then we might expect the random simulation to produce rather poorer data (i.e. equivalence classes with greater membership numbers to reach 90% coverage of the lexicon). Figure 2 superimposes the data from this second investigation on the data from Figure 1.

Figure 2. Distribution of equivalence classes for transcriptions which are MidClass throughout, MidClass with FineClass in stressed syllables, and MidClass with FineClass at random. Figures in parentheses denote mean size of equivalence classes.

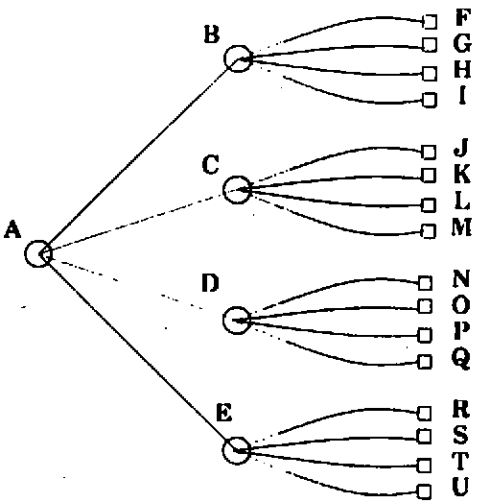



The only difference between the two FineClass conditions in Figure 2 is that in one case, the FineClass information is located only within stressed syllables, and in the

other, the FineClass information is located at random. In both cases, the same number of segments were affected: 33% of the segments were transcribed to FineClass, and 67% to MidClass. Despite the fact that the same number of segments were affected in both cases, we see a clear advantage when the FineClass information is located randomly throughout the lexicon, than when located systematically in stressed syllables: 91.5% of the lexicon is uniquely identifiable, and 99% using equivalence classes up to and including 5 members (mean class size = 1.06)

At first sight, this result is perhaps surprising. We might have expected the random condition to do as well as, or worse, than the stressed syllable condition. It turns out, however, that there is a simple explanation for this pattern of results. The essential difference between the two conditions is that in the stressed syllable case, the FineClass information lies on consecutive segments. In the random condition this is not the case. Consider the permissible sequences generated by the fragment of a finite state grammar in Figure 3.

Figure 3.




In this limited fragment, the probability of the event F given the event B is 0.25. The probability of the event F given the event A, however, is 0.0625: it is the probability of F given B multiplied by the probability of B given A. In general, the probability of a particular phoneme at position 1 given the existence of a particular

phoneme at position i-1 will be higher than the probability of the phoneme at position i given the existence of a particular phoneme at position i-n, where n > 1. In equivalence class terms, if the probability of one phoneme given another in the sequence is high, then we can expect those two phonemes to co-occur several times; we can expect several members in the equivalence class defined by that co-occurence. But if the probability of one phoneme given another is low, then we can expect those two phonemes to co-occur only rarely; we can expect only very few members in the equivalence class.

In a third simulation, in which every third segment was written to FineClass (i.e. 33% of the segments, but not at random), the overall distribution of equivalence classes lay somewhere between the random condition and the stressed syllable condition.

So FineClass information scattered at random is actually *more* informative than FineClass information over consecutive segments.

In the following section, we reconsider Huttenlocher [2], and the implications of these results for automatic speech recognition.

## DISCUSSION

The simulations described in the preceding section all demonstrate the power of using intermediary levels of description which lie between Broad and FineClass descriptors. The power of mixed class descriptions is also demonstrated. But the first simulation, in which only stressed syllables were transcribed to the FineClass level, questioned the degree to which stressed syllables are most informative. The results from the second simulation, in which *random* segments were transcribed to the FineClass level, would be expected *regardless* of the informative status of stressed syllables, and so do not shed any light on the validity of Huttenlocher's claims. A number of attempts to actually replicate Huttenlocher's findings, using the same 6 BroadClass units used by Huttenlocher, have in fact failed. In one instance only the disyllabic words in ALD were used: accounting for 39% of the lexicon (approximately 7200 words), and for which equal numbers of segments fell in stressed syllables as in unstressed syllables. No differences were found in terms of the distribution of equivalence classes when using Huttenlocher's procedure for deleting information contained in either the stressed or unstressed syllable.

So what are the implications for automatic speech recognition? There are a number of reasons for treating any statistics from lexicons such as ALD or Webster's with caution:

(1) The lexicon does not distinguish between words which are frequent in the language and words which are not. The implications of this are clear if we consider the hypothetical case where all the words which are uniquely identifiable are all extremely rare, and where the equivalence class which contains most members also contains the most frequently occuring words in the language. Ideally, the statistics should be collected from a lexicon which is weighted for frequency. Huttenlocher does in fact report results which are weighted for frequency. He states, however, that similar patterns of results were observed using an unweighted lexicon.

(2) The lexicon will not necessarily contain all the inflections of a word, and is not, therefore, representative of the language.

(3) The phonological transcription of each word will reflect its citation form. None of the many reduced forms will necessarily be included. Indeed, there may be many reductions associated with one citation form. Ideally, the statistics should be collected from a lexicon which includes such reduced forms. Again, the lexicon does not reflect the target language.

(4) Within the context of continuous speech recognition, even if each word in the input string were uniquely identifiable on a word-by-word basis from MidClass information alone, as soon as the words are strung together one is faced with the word boundary problem: during the process of lexical look-up, each segment potentially starts a new word (though this can be constrained to a certain degree by stipulating that a segment can only mark the beginning of a new word if the preceding segment could mark the end of a word). If that segment is transcribed to the level of FineClasses, then only words which begin with that FineClass segment can be hypothesised. If the segment had been transcribed to MidClass, then words beginning with that MidClass can be hypothesised. Clearly, there will be more of the latter. There are several other effects of MidClass information on lexical look-up; for instance, although a word might be uniquely identifiable at the MidClass level in some lexicon, fewer branches of the lexicon would need to be explored if it were transcribed to FineClass level than if it were transcribed to the MidClass level.

So even if each word in a dictionary such as ALD were uniquely identifiable according to statistics of the kind reported here, it is by no means necessary that a string of such words will itself be uniquely identifiable. But despite these caveats, it still remains the case that successful recognition is more likely the greater the lexical discriminability of the input words. Moreover, working within the context of continuous speech recognition does also carry certain advantages. For instance, syntactic information can be used to reduce the search space (i.e. to eliminate members of an equivalence class). Figures 1 and 2 would look very different if membership of an equivalence class had been restricted to words which each shared the same syntactic form class.

## CONCLUSION

We began by considering some further explorations of Huttenlocher's findings concerning the informativeness of stressed syllables. Given an acoustic front-end which might reliably report FineClass distinctions only on occasion, it would be desirable for the *more informative* stretches of the signal to be reported at this level. On the basis of Huttenlocher's claim, these stretches should correspond to the stressed syllables. The lexical statistics demonstrate, however, that stressed syllables are not maximally informative in the sense suggested by Huttenlocher. Indeed, the results demonstrate the desirability of a front-end which outputs occasional FineClass information not systematically, but at *random*.

## ACKNOWLEDGEMENTS

My thanks to Richard Shillcock for encouraging me to think about stressed syllables, to Jonathan Harrington and Gordon Watson for the syllabification rules, and to the Edinburgh University Speech Input Project at CSTR for providing additional support.

## REFERENCES

[1] J. Dalby, J. Laver, and S. Hiller, "Mid-class phonetic analysis for a continuous speech recognition system", Proceedings of the Institute of Acoustics, Vol. 8, (1986).

[2] D.P. Huttenlocher, "Acoustic-phonetic and lexical constraints in word recognition using partial phonetic information", S.M. thesis, M.I.T., Cambridge, Mass. (1984).

[3] D.P. Huttenlocher and V.W. Zue, "A model of lexical access based on partial phonetic information", Proc. ICAASP-84, Vol. 26, No. 4, 1-4, (1984).

Proceedings of The Institute of Acoustics

# Filtering of Phonetic Hypotheses in a Speech Recognition System using Grammatical Tag Transitions.

John C. Foster, James R. Hurford.

Centre for Speech Technology Research, University of Edinburgh

## 1. LEXICAL ACCESS OUTPUT -- INPUT TO SYNTACTIC COMPONENT

This paper describes work in progress on the syntactic component of a large-scale domain-independent machine-assisted speech transcription system. In the design of this system, a spoken utterance is analyzed by a phonetic/phonological front end, which presents a lattice of phonemes to a lexical access component, equipped with a lexicon currently containing 3962 words. The syntactic component takes output from the phonetic and lexical front end, in the form of a list of wordstrings. The words in these strings are all from the system's lexicon. The size of the list output from lexical access depends on (a) the accuracy of the phonetic front end processing in producing the desired string of phonemes, and (b) the possibility of alternative lexical analyses of the phoneme strings coming from the front end. Both factors are potent sources of data explosion.

The following are examples of wordstrings received from the lexical front end.

(1) This assumes completely accurate phonetic front end processing (which in fact never happens in our system, because the front end at present aims for accuracy only to the nearest phonetic midclass).

a. which tea party did judge baker go to
b. witch tea party did judge baker go to

c. which t party did judge baker go to
d. witch t party did judge baker go to

e. which tea party did judge baker go too
f. witch tea party did judge baker go too

g. which t party did judge baker go too
h. witch t party did judge baker go too

i. which tea party did judge baker go two
j. witch tea party did judge baker go two

k. which t party did judge baker go two
l. witch t party did judge baker go two

As can be seen from this example, the presence of homophones causes multiple output from Lexical Access.

(2) (With phonetic confusion of voiceless fricatives and front vowels.)

a. three chefs face a thief
b. free chefs face a thief
c. free chefs fierce earth if
d. three chefs fierce earth if
:
:
j. fresh if sphere sir thief
:
:
n. three chefs fierce if f

Similarly, then, phonetic confusion causes multiple output from Lexical Access. For a three-word input utterance, and given the current lexicon, the number of strings output from Lexical Access to the syntax component may be as high as 100,000. The problem would be worse with an expanded lexicon, and basing the lexicon on smaller units, such as morphs rather than words, would also tend to increase the number of strings output by Lexical Access.

In the current state of the model, there are no scores on wordstrings output from Lexical Access, and so no preferences on lexical grounds for particular wordstrings are expressed. The inclusion in future of scores on words reflecting (a) confidence rating from the front end, (b) word frequency, and (c) collocational information, would allow the output from Lexical Access to be ranked before the syntactic scoring applied. At present the syntax component takes an unranked input.

## 2. TAG LOOKUP

The data explosion doesn't end with the output from Lexical Access. Syntax adds its own multipliers to the explosion in the form of tags since a word may have several different syntactic tags associated with it.

A tag is a part-of-speech characterization. For example consider (3) below:

(3) (N (Common +)(Num sg))
(V (Vform bas)(Vtype -*np*))
(Aj (Degree bas)(Pos at*pred*))

The first tag here represents the information "singular common noun", the second the information "base form of a verb which may be either transitive or intransitive", and the third represents "adjective which is neither comparative nor superlative, and which can occur in either attributive or predicative position". The information needed to construct tags for words is

contained in the lexicon, but tags are not the same as lexical entries. They do not necessarily contain all the grammatical information about a word available from its lexical entry. the example in (3) shows the fullest possible information for the relevant tags given the current lexicon, but another possibility is for tags to consist only of the major category with no feature information given, as shown in (4).

(4) (N)
(V)
(Aj)

Tags, therefore, though derived from lexical entries, are structured relative to particular strategies adopted experimentally during the development of the syntactic filter. We have experimented both with highly informative tags as in (3) above and with severely stripped down tags as in (4).

Another difference between tags and the grammatical information in lexical entries should be mentioned. This relates to the asterisks seen in (3), as in "(Vtype -*np*)" and "(Pos at*pred*)". Actual lexical entries from which tags containing such expressions are derived could be, for example:

(5) (bear (N (Common +)(Num sg))
(V (Vform bas)(Vtype - np)) )

(free (Aj (Degree bas)(Pos at pred))
(V (Vform bas)(Vtype - np)) )

Ambiguities of two kinds are apparent here. Firstly, each word belongs to two major categories: *bear* is either noun or verb, and *free* is either adjective or verb. Secondly, within some major categories there are subcategories, and a word may belong to several subcategories. Thus both *bear* and *free* as verbs are either intransitive or transitive (in our terms), which is indicated by the feature "Vtype" being followed by TWO values, "-" and "np". Similarly, *free* as an adjective is either attributive or predicative, which is indicated by the feature "Pos" having the two values "at" and "pred". In many cases subcategorizations such as these reflect the possible immediate environments in which words may occur. Thus "(Pos at)" for an adjective reflects the fact that the adjective in question can occur immediately before a noun, and "(Vtype np)" means that a verb can occur immediately before a noun phrase. Many words are very versatile in their distributions and belong to several such subcategories, which is why they receive multiple values for such features in the lexicon.

In the exercise carried out to obtain statistics for tag transitions from a corpus the facility was provided for not constructing separate tags for every such grammatical subcategory represented in the lexicon. In principle, it

would be possible to assign any of four possible tags to *free* , for example:

(6) (Aj (Degree bas)(Pos at))
(Aj (Degree bas)(Pos pred))
(V (Vform bas)(Vtype -))
(V (Vform bas)(Vtype np))

Where features in lexical entries have several values, the facility exists to create a "portmanteau tag" conflating into a single expression the values of the feature. Thus, instead of all four tags in (6), the two tags in (7) can be created.

(7) (Aj (Degree bas)(Pos at*pred*))
(V (Vform bas)(Vtype -*np*))

The asterisks in these tags can be thought of as glue bonding several feature values into one composite value.

The machinery for creating tags out of lexical entries is controlled by setting variables for one's choice of (a) which features, if any, to include in tags, and (b) which multiple feature values to squash into composite values.

Depending on the choices made in creating tags from lexical entries, a word in a wordstring output from Lexical Access may receive more or fewer distinct tags at the tag lookup stage. Even with the barest tags, just representing a word's major category (e.g. N, V, or Aj), approximately one third of the words in the current lexicon receive more than one tag. Thus *bear* is looked up as both (N) and (V) tags. Other words receive more than two tags. We have experimented with several possible sets of tags. The output wordstrings from Lexical Access are translated into word-tag strings, a one-to-many translation, because of the factors just discussed. So the single wordstring *free bear skins* would generate 8 word-tag strings as follows:

(8)

| free | bear | skins |
|---|---|---|
| Aj | N | N |
| Aj | N | V |
| Aj | V | N |
| Aj | V | V |
| V | N | N |
| V | N | V |
| V | V | N |
| V | V | V |

The conversion to word-tag strings may increase the number of strings by a factor of up to 30, for a 7-word input utterance. As with homophony, there is nothing we can do to reduce the grammatical multifunctioning of words, but it is planned to design the system so as to generate preferred word-tag strings first, given ranking of the output from Lexical Access and interaction between the calculation of tag-transitions (see next section) and tag lookup. The mechanism for expressing a preference over word-tag strings is described next.

## 3. PROBTAGMATRIX

A set of binary transitional probabilities between grammatical tags, extracted from a tagged version of sector H (about 65000 words) of the Lancaster/Oslo/Bergen Corpus of Modern British English [1] is represented as the value of a global variable PROBTAGMATRIX. A sample extract from PROBTAGMATRIX is given below.

(9)

| | | |
|---|---|---|
| (Av) | (V) | .2383499 |
| | (P) | .1466769 |
| | (Aj) | .05500382 |
| | (Conj) | .05118411 |
| | ⋮ | |
| (Conj) | (N) | .08510638 |
| | (Pro) | .09606706 |
| | (V) | .08510638 |
| | ⋮ | |

So, for example, the probability of finding a verb (V) after an adverb (Av) is relatively high (.2383499), whereas the probability of finding a conjunction after a verb is relatively low (.05118411).

The system analyzes each word-tag string on the basis of these probabilities and a function TAGNEWSCORE, described below.

#### 4. RANKING THE OUTPUT

The function TAGNEWSCORE keeps a running average of the transitional probabilities between tags as the word-tag string is parsed left-to-right. Take a word-tag string such as the following:

(10) free(Aj) bear(N) skins(V)

Two tag transitions are involved here, Aj to N, and N to V. Looking these up in PROBTAGMATRIX, transitional probabilities can be assigned as follows:

(11) free(Aj) bear(N) skins(V)
.446281 .140564

An alternative word-tag string over the same wordstring would receive different transition scores, as below:

(12) free(V) bear(V) skins(V)
.2372673 .2372673

A score for the whole word-tag string is at present obtained by averaging the transitional probabilities across the whole string. So the score for the word-tag string in (11) is 0.2934225, whereas the score for the example in (12) is 0.2372673. Thus (11) "beats" (12). We aim to investigate more sophisticated functions than simple averaging.

As mentioned, a single wordstring may correspond to several word-tag strings. The score assigned to a wordstring is the highest score assigned to any of its word-tag strings. By this means, a ranking of the wordstrings input from Lexical Access is obtained.

## FILTERING OF PHONETIC HYPOTHESES

### 5. VARIATIONS AND PRELIMINARY RESULTS

We have experimented with various alternative details within the framework outlined above. The principal variations developed to the point of usability within the system are: a larger tagset, word frequencies and trigrams The larger tagset, which for the purposes of the present experiments was set to 190 members, is illustrated in (3) above. Essentially, the difference between the larger and the smaller tagset(which for the purposes of this excercise has been set to 29 members) is that the former contains feature information in addition to the major category to which a word belongs. We suspect that some middle-sized tagset will prove better than either our larger or our smaller tagset as a tool for filtering out undesirable word-strings.

We will not discuss the use of individual word frequencies or trigrams in this paper.

Figures 2 to 4 summarize the main results of a series of tests run on the system using simulated and selective phoneme confusions and 15 of the test sentences chosen by the phoneticians working on the front end. Note that these test sentences are not from the same source (sector H of the LOB corpus) as the matrix of transitional probabilities used in the system.

The procedure for the tests was as follows. "Confuser-parser" software was set up to accept the 15 test sentences and simulate the performance of a phonetic front end in producing a number of different hypotheses as to what phonemes the input utterance might consist of. The assumption was that the phonetic front end would confuse phonemes with similar manner but different place of articulation. So, for example, the three nasal consonants (m, n, ng) could be confused or the three voiceless stops. Figure 1 identifies the meanings of the 13 classes of phoneme confusions used in these experiments. A summary of the confused strings ranked according to the probabilistic scores generated by the syntactic component was stored. No Front End information in the form of individual word or phoneme hypothesis scores was used in generating these syntactic scores.

Figure 2 is a summary of the average number of confusions generated by the Lexical Access component on the assumption that only one class of phonemes(e.g. voiceless fricatives, or front vowels) was confused. It can be clearly seen that even with no confusions (see column labelled "NIL"), homophones in the lexicon cause nearly eight strings to be offered to the syntactic filter for parsing. This number rises considerably with nasal confusions (column labelled "N") and diphthongs ("D").

Figure 3 tabulates the average rank orders of the input test sentence for the large and small tagsets and each of the confusions. The small tagset did slightly better than the large tagset although the difference in performance was not statistically significant.

Figure 4 tabulates the percentage success of each of the tagsets with respect to specific rank orders. Again, it is clearly seen that the small tagset did slightly better than the large tagset.

In conclusion, it appears that the technique of ranking word strings on the basis of transitional probabilities between adjacent grammatical tags is potentially useful in a speech recognition system.

## 6. REFERENCES

[1] Johansson, S., G. N. Leech, H. Goodluck. Manual of information to Accompany the Lancaster-Oslo Bergen Corpus of British English for Use with Digital Computers, Department of English, University of Oslo, 1978.

| | |
|---|---|
| P | Voiceless stops |
| B | Voiced stops |
| S | Strong voiceless fricatives |
| F | Weak voiceless fricatives |
| Z | Stong voiced fricatives |
| V | Weak voiced fricatives |
| N | Nasals |
| L | Liquids |
| G | Glides |
| FV | Front vowels |
| CV | Central vowels |
| BV | Back vowels |
| D | Diphthongs |

**Figure 1: Phonetic Values for Abbreviations used in the Experiments.**

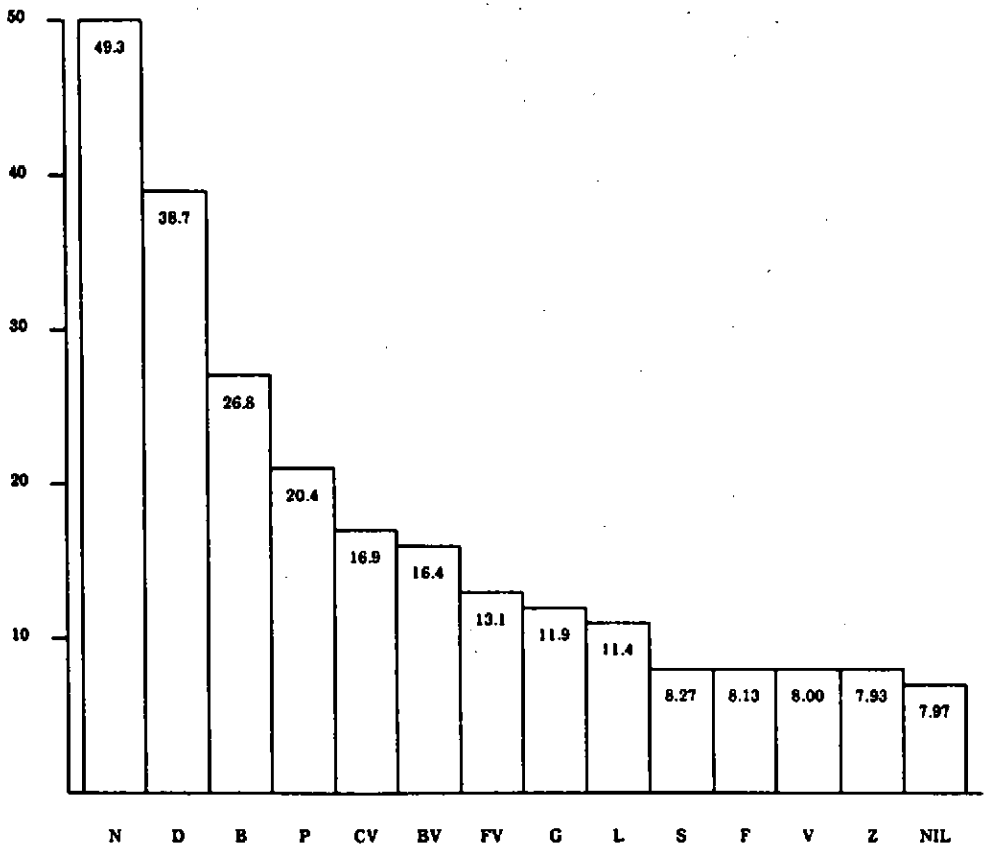



**Figure 2** the average number of strings generated across the phoneme lattice for the 15 test sentences and a single mid-class confusion

| tagset | N | D | B | P | CV | BV | FV | G | L | S | F | V | Z | NIL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| large | 2.6 | 2.73 | 5.15 | 1.27 | 1.33 | 1.27 | 2.0 | 1.29 | 1.33 | 1.27 | 1.2 | 1.2 | 1.6 | 1.13 |
| small | 1.87 | 1.33 | 1.67 | 1.73 | 1.27 | 1.33 | 1.33 | 1.27 | 1.33 | 1.33 | 1.33 | 1.27 | 1.27 | 1.27 |

**Figure 3 Average rank order with each tagset**

| rank order | large tagset | small tagset |
|---|---|---|
| 1 | 74.8 | 82.9 |
| 1 - 2 | 93.8 | 90.5 |
| 1 - 3 | 97.1 | 91.4 |
| 1 - 4 | 97.1 | 97.6 |
| 1 - 5 | 97.1 | 98.6 |

**Figure 4 Percentage success of the two tagsets in achieving specific rank orders**