UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Word Recognition: A Paradigm Case for Computational (Psycho-) Linguistics

Permalink
https://escholarship.org/uc/item/8fq4q3r9

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 6(0)

Author
Thomson, Henry

Publication Date
1984

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Word Recognition:

A Paradigm Case for Computational (Psycho-) Linguistics[1]

Henry Thompson
Department of Artificial Intelligence
and
Program in Cognitive Science
University of Edinburgh

1. Introduction

Existing psycholinguistic models of word recognition are incapable of computational realisation as currently formulated. Existing computational linguistic approaches to lexical access, morphology and spelling correction provide a rich repetoire of computational methods which appear to offer attractive possibilities for addressing this problem, but are not as they stand psychologically plausible. The proposals made in this paper come out of the efforts of an interdisciplinary group in the Programme in Cognitive Science of the School of Epistemics at the University of Edinburgh to reconcile these two facts, and arrive at a psycholinguistic model of word recognition in continuous speech which is explicitly and realisably computational.

The focus of this work is on the normal processing of continuous speech, and we take three points to follow necessarily from this:

1. Word boundaries are not unequivocally marked

2. The speech signal is not always in and of itself sufficient for recognition,

   and in particular

3. The initial segments of words are not acoustically reliable

Two possible sources can in principle be identified for the solution to the problem posed to the hearer by the second point above: Linguistic structural knowledge (e.g. the phonotactics, lexicon and grammar of the language involved) and contextual/pragmatic knowledge. Both computional experience (most notably in the ARPA speech understanding research effort in the 70's) and psycholinguistic experiments (e.g. (Pollack and Pickett 1963), (Lieberman 1963)) have given strong support to the position that the information required by these sources to recognise the speech signal sometimes occurs temporally after the under-determined section - the so-called right context effect.

Despite this the model we propose, which is for the moment concerned with only the deployment of lexical and morphophonemic structural knowledge, is strictly left-to-right and bottom-up. In the next sections a sketch of the

---

[1]Many of the ideas presented here were devloped in discussions with Ellen Bard, Gerry Altmann and Anne Johnstone, for whose constructive criticism I am grateful. My debt on the computational linguistics side to Ron Kaplan and Martin Kay (Kaplan and Kay 1982), for their ideas and support, is substantial.

model is presented in sufficient detail to show how it can none-the-less generate the right context effect.

The focus in these sections is on the computational characterisation of the model - owing to space restrictions the psycholinguistic consequences will largely be left implicit.

The basic question to be addressed is how to make effective computational use of the dramatic restriction on the problem of speech recognition which is provided by the fact that the signal to be analysed is constrained to consist (by and large) of sequences of (for the sake of argument) English words. Two knowledge bases, lexical and morphophonemic, are required, plus a computational method for deploying them.

The lexical knowledge base consists of a set of trees in which morphemes (stems or affixes) sharing the same initial phoneme sequence follow the same path until their phonemic representations diverge. Recognising a word then consists of stitching together a morphotactically valid path through these trees, and recognising a whole utterance consists of concatenating a sequence of such paths.

The morphophonemic knowledge base consists of a set of (possibly context-sensative) rewrite rules, which are expressed as a finite-state transducer which is interpolated between the lexicons and the input.

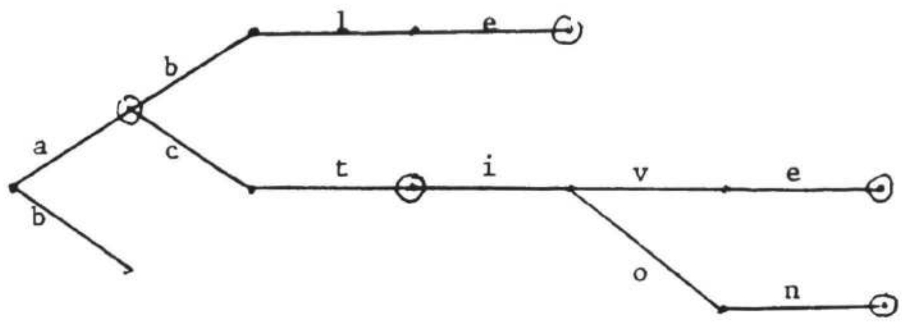
The easiest way to understand this approach is to start with a version appropriate to text, and then suggest how to modify it for speech.

2. The Recogniser - Text version

2.1. The Lexicon

In its simplest form, with only one tree-structured lexicon, looking up a word is straight-forward. One starts at the top of the tree and follows the branch whose label is the first letter of the word in question. From the end of that branch, one follows the branch whose label is the second letter of the word, and so on until one either runs out of letters or branches. In the first case, provided there is an indication in the tree of a word ending at this point, the process has completed satisfactorily. Otherwise, either the dictionary or the candidate word is faulty.

A trivial example will clarify this. Consider the following (extract from a) tree, where circled nodes indicate the presence of words, with lexical entries attached:



If one looks up 'a', 'act', 'able', 'active' or 'action', the process will succeed, but if one tries 'abd', 'abled' or 'acti' the process will fail in one way or another.

Although this approach is clearly much more efficient than one which compares candidate words to entries in a sequential lexicon one after another, it would seem of no interest once we move from text to speech, as it ignores point (1) above, depending as it does on prior knowledge of just that thing we have said is not available in speech - an indication of where words begin. But by elaborating two features already alluded to above, we can then confront this problem in an interesting way.

Firstly, the system can be made to process not words, but strings of characters. Its goal is to find a sequence of paths through the tree which, when laid end to end, will cover the entire target string. To accomplish this it treats spaces and punctuation as alphabetic characters, including them in the lexicon as well.

Secondly, the lexicon consists not of one tree, but of a collection of trees or sub-lexicons, one for each morphological class, including prefixes, stems, noun inflections, verb inflections, derivational suffixes, punctuation and so on. Morphotactic constraints, insofar as they are finite-state, can be expressed by including in each lexical entry an indication of the sub-lexicon(s) where processing should carry on.

The crucial point about this approach is that spaces and other punctuation have now lost their special status with respect to determining word boundaries. It is the act of wrapping around back to the stem (or prefix) sub-lexicon that marks a word boundary.

2.2. Non-determinism and the Chart

This move, as well as other characteristics of the above proposal, introduces a non-deterministic element into the lookup process. A principled approach to this is required, as it will prove crucial once we move to speech. We take the approach of representing all the data in the system, at the character, morph and word level, as edges in a lattice or chart, and adopt the Active Chart Parsing methodology for discharging the non-determinism (Kay 1980), (Thompson and Ritchie 1984). This methodology supports inter alia a left-to-right pseudo-parallel process which guarantees that all hypotheses about possible paths will be pursued, and allows the search strategy through the space of such possibilities to be closely controlled. For example if "r e d" is an initial substring of a text to be analysed, provided the relevant lexical entries are present, three hypotheses will be current after processing those three letters. One will be for the word "red" followed by some as yet unseen inflection and/or punctuation; one for the set of English stems beginning "red", e.g. "rede", "redcoat", "redeem"; and one for the prefix "re-" followed by the first letter of a stem beginning with "d".

The chart-based approach allows ambiguities to be perspicuously and economically recorded. The letters "u n i o n i s e" for instance would be spanned by two edges, one for the stem "union" plus the suffix "-ise", and the other for the prefix "un-", the stem "ion" and the suffix "-ise".

2.3. Morphographemics

The chart also provides a convenient basis for dealing with morphographemics. One can either view morphographemic rules as rewriting rules to be applied by adding edges to the chart (thus y -> ie / _ s can be seen as calling for the addition of a pair of edges with a 'ys' on spanning any 'ies' trio of edges, which then allows e.g. "flies" to be correctly analysed as "fly" plus "-s") (Kay 1977), or as transducers to be interpolated either serially (Kaplan and Kay 1981) or in parallel (Koskenniemi 1983) between the chart and the lexicon.

3. The Recogniser - Speech version

The conversion of the preceding text-based system to handle speech is surprisingly straight-forward. The principal differences stem from the necessity to handle the imperfection of the 'input' acoustic data.

3.1. The Lexicon

Here the change is minimal, simply replacing letter-based branches with phoneme-based branches, and eliminating the punctuation sub-lexicon.

3.2. Non-determinism and the Chart

The chart provides a convenient device for dealing with the indeterminacy of the acoustic data. As in the ARPA speech projects we represent that indeterminacy with a lattice of alternative phonemic analyses, with each alternative analysis of a given segment represented by a distinct edge in the chart.

The Active Chart Parsing methodology insures that all possible matches between phoneme edges and lexicon branches will be tried, but this is not in itself enough. Matching of chart edges against lexicon branches can no longer be considered all or nothing. The imperfections of the input data most be accounted for at this point. For the sake of exposition we will assume that the phoneme edges have some score between 0 and 1 associated with them, representing the degree of certainty assigned to them by the acoustic-phonetic levels of the system.

It follows that hypotheses about the presence of morphs, both complete and partial, will have scores associated with them as well. In the first instance we can think of computing the score incrementally, as a partial hypothesis is extended to cover another phoneme. The score of the new hypothesis will be some function of the score of the prior hypothesis, the score of the newly incorporated phoneme, and the goodness of match between that phoneme and the branch followed in the lexicon. For example if a partial hypothesis existed analysing the initial two segments of a signal as /r e/, with score 0.7, and one of the segments in third place was a /b/ with score 0.8, then we would get among others two new hypotheses, one for /r e b/ with score say 0.75, and one for /r e d/ with score say 0.6.

It follows that there will be a vast increase in the number of partial hypotheses entertained at any given point in the processing. The scoring is what provides a mechanism for keeping this explosion under control. We suppose that at any given point only some number of the highest scoring hypotheses are extended to yield new ones. It is this which produces the appearance of the right context effect. Suppose the actual word uttered in the preceding example had been "reduplicate". In the process of extending the two hypotheses the successors to the one beginning /r e b/ will slowly drop in score, as there is no lexical path which will match well against the input, while the successors to the one beginning /r e d/ will slowly climb in score along the correct path, in effect retroactively confirming the /d/ hypothesis and rejecting the /b/ hypothesis.

An additional level of complexity is required to deal with the issue of word boundaries. A complete word hypothesis is not independent of the subsequent processing it entails. Going back to our example, any hypothesis that the sample in fact begins with the word "red" must be down-graded by the necessarily low score of any hypothesis of a word beginning /oo p l/. Whether this is accomplished by filtering at a higher level, by coupling the scores of successive hypotheses together in some way, or by recording complete hypotheses only after successfully (to some level of score) identifying several words in a row, is a matter of some uncertainty at the moment.

3.3. Morphophonemics

The text-based approach translates easily into a speech-based one, retaining the three options described under that head for the interpretation of the rules.

4. Meta-theoretical Considerations

The model as presented here begs many questions, both down the speech chain, by assuming a phonemic segmentation, albeit imperfect, as input, and up the speech chain, by saying nothing about how syntactic and semantic effects may be felt. We believe the left-to-right, selective interaction approach can work at these levels as well, but are not yet in a position to offer details of how this might be done. The incorporation of many of the ideas presented here in a large scale speech-processing system, now underway, will give substantial impetus to providing such details in these areas as well.

The point we hope to have illustrated with even this brief sketch is that a careful investigation of the computational tools available may reveal a straight-forward way of providing an account based on a left-to-right process using only selective interaction of a phenomenon usually held to require a right-to-left process using instructive interaction.

The question which must at this point in the discussion be most pressing however is that of computational detail and psychological relevance. Surely a model which is so computationally explicit makes far too many detailed predictions to withstand experimental test. This may prove to be the case, but we feel that if Cognitive Science is to be more than just an indulgent pastime for psycholinguists, then the attempt at fully explicit computational models must be made, so that the consequences in terms of submitting such models to experimental verification can be made in practice, rather than in theory. We furthermore feel that the closer one can get to the automatic periphery of the perceptual system, the better chance one has of making a useful experimental contribution to the development of computational models, and vice versa. In this we see ourselves as in a small way trying to import into the linguistic modality the methodology which was so sucessfully introduced by David Marr for the visual modality.

REFERENCES

Kaplan, R.M. and Kay, M. 1981. Phonological rules as finite state transducers. Presented at the Winter Meeting of the Linguistic Society of America, New York.

Kaplan, R.M. and Kay, M. 1982. Word recognition. Technical Report, Xerox Palo Alto Research Center. To appear.

Kay, M. 1977. Morphological and syntactic analysis. In Zampolli, A., editor, Linguistic Structures Processing. North Holland.

Kay, M. 1980. Algorithm Schemata and Data Structures in Syntactic Processing. In Proceedings of the Symposium on Text Processing. Nobel Academy. To appear. Also CSL-80-12, Xerox PARC, Palo Alto, CA.

Koskenniemi, K. 1983. Two-Level Model for Morphological Analysis. In Bundy, A., editor, Proceedings of the Eighth International Joint Conference on Artificial Intelligence. IJCAI, Los Altos, CA.

Lieberman, P. 1963. Some effects of semantic and grammatical context on the production and perception of speech. Language and Speech 6:172.

Pollack, I. and Pickett, J.M. 1963. The intelligibility of excerpts from conversation. Language and Speech 6:165-171.

Thompson, H.S. and Ritchie, G.D. 1984. Techniques for Parsing Natural Language: Two Examples. In Eisenstadt, M., and O'Shea, T., editors, Artificial Intelligence Skills. Harper and Row, London. Also DAI Research Paper 183, Dept. of Artificial Intelligence, Univ. of Edinburgh.