DURATIONAL SHORTENING AND ANAPHORIC REFERENCE

W. N. Campbell

ATR Interpreting Telephony Research Laboratories, Kyoto, Japan.

ABSTRACT

A tendency for the durational shortening over time of nominal heads and pronominal references to them was confirmed in a twenty-minute radio broadcast of a short story. Lengthening was calculated by comparing durations predicted by the timing algorithm of a computer text-to-speech program with those measured from a recording of the story after factoring out global changes in rate of speech. Considerable variation was noted in the residuals thus formed, and resetting was found to correlate with events in the narrative of the text.

1. INTRODUCTION

Fowler & Housoum [4] have shown that speakers distinguish words that are *new* to a monologue from those that are assumed *given*, by shortening subsequent occurrences of the latter. More recently, for Dutch, Eefting [3] constructed experiments to compare the effects of *information value* and *accentedness* on the duration of words but found only a weak effect for the former, compared to that of the latter, and questioned the value of given-ness as a predictor of duration in itself.

This paper looks at the durational correlates of anaphoric reference in a long passage of naturally occurring professionally narrated text and shows that the durations of both antecedent nominal units and subsequent pronominal references to them can be seen to reduce as the utterance progresses, but with resetting occurring after an interval of time and at major breaks in the narrative of the text.

2. MATERIALS

A twenty-minute radio broadcast of a short story was analysed for anaphoric reference, enumerating all nominal units and tagging pronominal references to each with an identifying code. These codes were linked using a text editor to a list of the syllables with measures of the lengthening undergone by each. In this way every occurrence of a nominal unit or any of its referring pronouns can be associated with a value representing the degree of reduction in the duration of its syllables.

2.1. Measuring length

The passage was digitised and measured for duration at the syllable level. The timing component of a computer text-to-speech system was optimised for the passage and used to predict the durations from input that had been manually coded to describe each syllable in terms known to control a large portion of the durational variance. The output from the program was compared with the original durations, and a set of residuals produced (*residual* (%) = (*predicted duration/observed duration* x 100) – 100) which can be taken to represent the amount of over- or under-prediction, which in turn can offer clues to the identity of significant factors not being suficiently taken into account by the prediction algorithm [2].

The strongest of these factors is presumed to be variation in speech rate, which is clearly evident on perceptual evaluation and can be visualised in a graphical plot of the residuals as a slowly

changing, low frequency offset. The *supsmu* smoothing function of the Splus statistical package [1] was used to model this offset and the smoothed representation then subtracted from the residuals to factor out, in a simple way, that part of the variance that can be assumed due to changes in speaking rate. The new residuals thus obtained indicate whether the speaker was rendering each syllable faster or more slowly than the local norm. If the program predicts a duration greater than that observed, then the residual will be positive, reflecting a presumed reduction in the duration of the spoken syllable, and vice versa.

Such data are by no means perfect, and results for individual samples are subject to considerable uncertainty due to measurement errors etc., but if trends can be observed from large numbers of observations then appropriate rules to describe the trends can be formulated and incorporated into the model to improve the quality of future predictions. Some loss of certainty is inevitable when working with large numbers of samples, and the method lacks the controls that experiments with laboratory-produced sentences may have, but this is felt to be a small price to pay for insights into the more delicate timing processes of naturally-occurring speech data.

3. RESULTS

Seven nominal units were repeated more than ten times in the passage:

nominal:	freq:	references:
Gerry	18	he 174, him 44, you 12, I 4
the tunnel	14	it 4, there 1
the rock	14	it 5
the water	14	- -
his mother	12	she 29, her 17
his head	12	it 1
the boys	11	they 16, them 11

No unit was monosyllabic, and some references contained adjectives or were homonyms, as in *a well of blue sea, the stinging salt water, the blue well of water, etc.* for *the water*. In such cases the mean value of lengthening for all syllables in the group was taken to represent the lengthening of the unit.

The duration prediction algorithm accounted for 86% of the variance, with a correlation of $r = 0.93$. After deduction of the smoothed fit, the standard deviation of the residuals was reduced from 20ms to 7ms, and the range reduced from 138ms to 58ms. Values of lengthening were in the range of ±20%.

In very few cases was there a simple linear reduction in duration over time. The overall trend was towards a reduction in syllable duration, but there was considerable oscillation about the reducing mean. With the exception of the hero's name, *Gerry*, for which the slope was -0.12, robust regression lines fitted to the samples showed a positive slope, the steepness of which (0 = flat = no change) indicates the rate at which later syllables are reduced in duration.

If we look, for example, at references to *the rock* (Figure 1), a slope of 1.29 fits in the overall case, but there is a major change of direction in the narrative after ref. 4, and a new sequence can be considered to begin from ref. 5. The resulting slopes for the two groups thus formed would become 7.8 and 3.6 respectively.

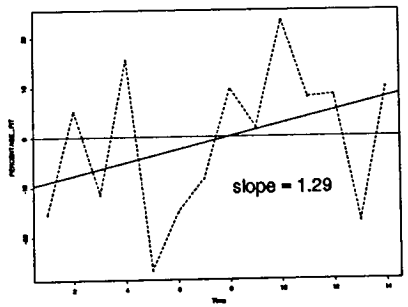

Figure 1: Residuals for *rock*.

For occurrences of *the water* (Figure 2), the overall slope was positive at 0.035,

but not significant. Again there is a major break in the theme of the text after ref. 7, and a line fitted to the first 7 occurrences shows a steeper slope of 6.14, from the percentage values -16, -12, -1, 7, 5, 18, and 20. The next three occur closely in time at 6m 40s, 6m 50s, and 7m 40s and if a reset is assumed to have occurred between these and ref. 11, occurring much later at 10m 25s, corrected slopes for the next two groups thus distinguished become 10.83 and 1.41.

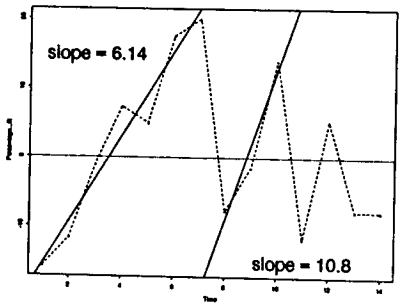

Figure 2: Residuals for *water*.

Time alone is not a sufficient cue to resetting; references to *the boys* show a similar pattern, and resetting after the first four would change an overall slope of -1.9 into a slope of 9.2 for the first part. However, although refs. 4 & 5 are very close together (at 5m 55s and 6m 10s), during these 15 seconds of narrative, Gerry goes swimming underwater, exploring, comes back to the shore, and then sees the boys once again in the same place as he first saw them. Not only is the location reset, but also apparently the syllable timing.

Pronomial reference too, shows similar reduction in duration with time. A robust regression line fitted through the 12 references to 'Gerry' as *you* (Figure 3) had a slope of 1.76 to fit the percentage values (-12, 22, 10, 7, 34, 24, 10, 22, 6, 24, 31, 24) measured for each occurrence.

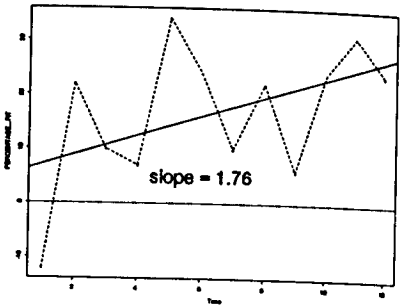

Figure 3: Residuals for *you*.

If the first six tokens, which occur together in the first 1m 50s of the story are considered separately from the next, which does not come until 11m 29s, steepness of the two groups becomes 6.2 and 3.2 respectively.

References to Gerry as *him* fit a slope of 0.6 (Figure 4). It seems that the first few group separately, and the last, too, may be an exception.

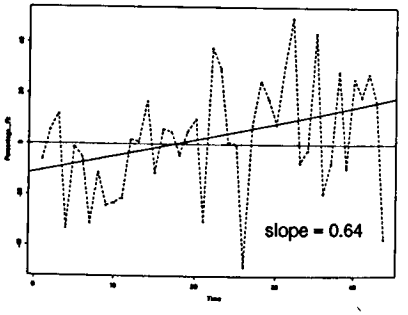

Figure 4: Residuals for *him*.

References to his mother as *her* (Figure 5) show less variation about the regression line, which has a slope of 2.79. The first three points here, too, appear exceptional; no explanation has yet been found for why these tokens should be spoken faster than the subsequent ones, but it is possible that the smoothing algorithm used to factor out global changes in speaking rate coped less well with the edge samples. However, such local details are beyond the necessarily general scope of this paper and would require more sophisticated statistical and linguistic analyses.

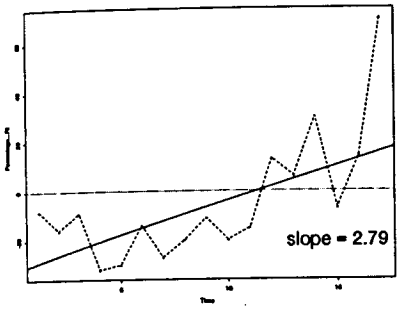

Figure 5: Residuals for *her*.

5. DISCUSSION

It would be unwise to draw strong conclusions from such a small study, but a clear trend has been shown, with regression lines drawn through selected groups of the data having a positive slope of around 5, indicating a 5% difference in the fit of each subsequent occurrence after factoring out the most obvious effects on duration by a predictive model. This notional line describes the best fit to the distribution of the points, but there is considerable scatter around the line and it would be difficult to predict the degree of fit for any one token from it. This difficulty is compounded by the difficulty in knowing where to allocate a reset. Clear separation in time appears to be one factor, and in retrospect, with the aid of the difference measures, marked changes in the flow of the narrative can be found that correlate well with resets in the cycle, but work has yet to be done to determine whether such resetting points could be determined *a priori*, from knowledge of the text alone.

On the other hand, the data do show that such work would be of use to the prediction of timing in speech, and confirm that there is a tendency to shorten the duration of items in a narrative when those items can be assumed to be known, or shared information. The samples were taken from a long passage of naturally occurring professionally narrated text and the measuring of their length made use of a differencing between observed and predicted durations using the timing algorithm developed for a computer text-to-speech system. This one-step-removed form of measurement introduces a degree of uncertainty of its own, but allows a finer view of the effects on speech timing by factoring out the grosser, more predictable components.

Acknowledgements

I am particularly grateful to Steve Fligelstone of Lancaster University for providing the Anaphora tagging, and to ATR for enabling the continuation of this research.

REFERENCES

[1] BECKER, R. A., CHAMBERS, J. M. & WILKS, R. A. (1988) **"The New S Language: A Programming Environment for Data Analysis and Graphics"**, AT&T Bell Laboratories, Wadsworth & Brooks/Cole Advanced books & Software, Pacific Grove California.

[2] CAMPBELL, W. N. (1990) *"Measuring Speech-Rate in the Spoken English Corpus"*, pp 61 - 81 in **Theory and Practice in Corpus Linguistics**, Eds J. AArts & W. Meijs, Rodopi, Amsterdam.

[3] EEFTING, W. (1991) *"The effect of 'information value' and "accentuation" on the duration of Dutch words, syllables, and segments"*. JASA # 89, pp 412 - 424.

[4] FOWLER, C. A. & HOUSOUM, J. (1987) *"Talkers' signalling of "new" and "old" words in speech and listeners' perception and use of the distinction"*. J. Mem. Lang. #26, pp 489 - 504.