INFLUENCES OF CONTEXTUAL PREDICTABILITY AND LEXICAL PROSODY ON ESTONIAN WORD DURATION

Liisi Piits & Meelis Mihkla

The Institute of the Estonian Language, Estonia
Liisi.Piits@eki.ee; Meelis.Mihkla@eki.ee

ABSTRACT

The article investigates how different factors such as word predictability and part of speech may affect word duration in Estonian speech. The material comes from corpora of read texts. On the example of the five most frequent words in the material (*eesti* 'Estonian', *ei* 'not', *ja* 'and', *on* 'is; are', *see* 'it; this') the correlation of the predictability and duration of words is studied. It is concluded that more frequent collocations are pronounced shorter, while the left collocate tends to be slightly more important for the node word duration than the right one. The modelling of speech temporal structure requires a specification of parts of speech as, depending on the part of speech, different factors (frequency, collocational strength etc.) have a different influence on the node word duration.

Keywords: collocational strength, word duration, lexical prosody, predictability, significant features

1. INTRODUCTION

One of the motivators for the present study was our observation of vacillations in the speech rate of radio newsreaders, however professional, when recording for the corpus designed for corpus-based synthesis of Estonian speech. The slowing down may have been due to more difficult sound clusters (the corpus was to contain all diphones possible in Estonian, including extremely rare ones [4]), which in turn occurred in very infrequent words. Speech rate enhancement, however, may be connected with frequent words as well as with collocations. For English, which is a language with strict word order, it has been proved that both the high frequency of occurrence and the predictability of a word may have a reducing effect on its pronunciation [2, 5]. One of our aims was to find out whether context may affect word duration in Estonian as well, although it is a synthetic language with a relatively free word order. The starting hypothesis was as follows: word duration is affected by collocational strength so that the words co-occurring more frequently are pronounced shorter.

Traditionally, studies of lexical prosody take a separate approach to content and function words e.g. [1, 3, 7]. As Estonian is a language with rich morphology, with no prepositions or articles, we got interested in a second hypothesis, asking what if models of the temporal structure of Estonian speech are not adequate enough if based just on the binary opposition of content vs function word, and the models should better be built on a more detailed division of parts of speech.

To test our hypotheses we used some statistical methods (regression and CART), believed to be able to detect some small, hidden, yet significant influences on word duration [6].

2. MATERIAL

The original material came from corpora of fluent speech containing longer passages of radio news (10 and 15 minutes of speech) and from speech synthesis corpora (51 and 66 minutes). The passages were measured for prosodic information, esp. word duration, and the left and right collocates of the node were found. The frequency of each word and the frequencies of its co-occurrences (collocation span: L1-R1) were computed from the journalistic subcorpus (5 million word forms) of the balanced corpus of the University of Tartu. The collocational strength was found using the following formula:

$$collocational\ strength = log\left(\frac{co\_occurence\ frequency}{nodeword\ frequency}\right) \quad (1)$$

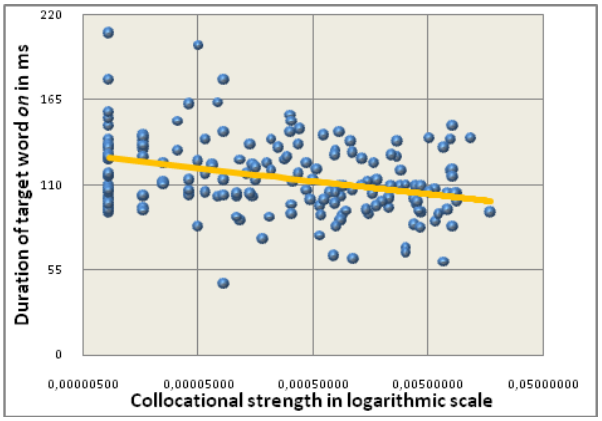
3. RESULTS

3.1. Correlation between word duration and collocational strength in frequent words

To test the hypothesis that Estonian, too, may display a correlation between the pronunciation length and contextual predictability of a word we first chose the verb *olema* 'be' as the most frequent

content word in Estonian, and computed the co-occurrence frequencies of all forms of *olema* found in the corpus with their right collocates. The verb *olema* was represented in the corpus in 23 different forms. To make them comparable the length of two stem phones only was measured. Next a simple durational model was built to predict the duration of the available forms of the *olema* verb from collocational strength, phrase length, position of the node verb in the phrase, and a binary variable indicating whether the concrete verb form is monosyllabic or not. According to the resulting models there were two significant parameters, the binary one and collocational strength. Consequently, at least for the Estonian verb *olema* 'be' a weak correlation can indeed be observed between collocational strength and word duration (see Figure 1).

Figure 1: Scatter diagram between the node word *on* 'is; are' duration and collocational strength with its right collocate.



In order to find out whether the discovered connection between the input and output was regular or occasional we had to increase both the speech material and the number of node words to be tested. Even increased amount of speech material was relatively small to give sufficient co-occurrence information for every word, so we had to focus on the most frequent words in the corpus, from which, in turn, we selected five word forms representing different parts of speech: *eesti* 'Estonian', *ei* 'no; not', *ja* 'and', *on* 'is; are', *see* 'this; it', found the collocates for each occurrence of the word forms selected, and calculated their collocational strength with their left and right collocates. Table 1 reveals that the correlation is systematically stronger concerning the left collocate. The other unexpected result is related to the function word *ja* 'and', notably, *ja* as the most frequent function word in Estonian displays no connection whatsoever between the collocational strength and duration. This may be due to the fact that the function word *ja* often lies on the boundary of a prosodic phrase. Thus there may be either a pause before *ja* or a phrase boundary marked by a lengthening of the word *ja*.

Table 1: Correlations between node word duration and collocational strength.

Node word	Correlations between node word duration and collocational strength	
	With left collocate	With right collocate
eesti	-0.17	-0.14
ei	-0.27	-0.10
ja	-0.07	0.00
on	-0.18	-0.15
see	-0.15	-0.11

3.2. Lexical prosody

In Estonian the word has a very important role both in grammar and in phonetics, while the language has an extremely rich morphology. Hence we developed an interest in whether morphological and lexical features could have any influence on the temporal structure of Estonian speech. The huge number of the possible word forms as well as the great variety in compound word formation impedes direct estimation of the effect of either part of speech or morphological characteristics on the duration of all word forms, however large the corpora. Therefore the effect had to be studied indirectly. To be more precise, we investigated the variation of phone duration as depending on the part of speech and the form of the word. Two methods, linear regression and CART, were used in modelling the correlations. For a qualitative estimation of the effect of each factor of interest the change in the output error of the models was measured and compared. The results showed that addition of morphological and part-of-speech information in the input meant a few percent decrease of the output error.

The most distinct regularities were revealed by visual assessment of the regression models of the part-of-speech factors. Table 3 gives the mean shortening/lengthening of phones in different parts of speech vs. the verb phones, separately for male and female readers. The grey background sets off the function words. As we can see there is more variety in the middle part of the table, while the parallel readings in its beginning and end are rather close. According to Table 2, in proper names the phones are pronounced about 5-6 ms longer, on

average, than in verbs. The average phone duration of the readers involved was 62.5 and 64.1 ms, which means that they took about 10% more time with proper names than with verbs. A little more time was also spent on nouns and adpositions. It was certainly surprising to find adposition phones to be longer than average. The adposition, after all, belongs to function words, which in most languages are pronounced shorter than the content words. In a sentence, a typical Estonian adposition goes with a noun (while that noun is often positioned in sentence focus), is pronounced longer than average, and may influence the context up to the next adposition. The ordinal numbers, however, were pronounced over 10% shorter, while a 5%-shortening was observed in pronouns and adverbs. The shortening of ordinal numbers can perhaps be explained by most of them denoting years: correct reading practice requires pronunciation of the whole long number, but if the century is the same, the reader soon tends to shorten the first half.

Table 2: Mean lengthening/shortening of phones (ms) in different parts of speech.

Part of speech	*Male speaker*	*Female speaker*
Proper name	6.23	5.22
Noun	2.25	2.10
Adposition	0.82	2.82
Genitive attribute	0.42	1.35
Verb	0.00	0.00
Numeral	–0.10	0.42
Conjunction	–0.14	1.81
Adjective	–0.39	1.14
Adverb	–0.89	–2.90
Pronoun	–4.13	–3.86
Ordinal numeral	–5.44	–7.48

According to the results, addition of morphological and part-of-speech information to the input of a durational model will reduce the prediction error by a few percent, which is really no surprise considering the essential role of the word as such in Estonian grammar as well as phonetics. Even though on average the function words were shorter than the content words, the relative shortness of ordinal numbers and the relative lengthening of adposition phones indicate that the binary division of words into content and function words is too crude a parameter to model Estonian lexical prosody.

3.3. Significance of features of predictability and lexicality for models of word duration

The building of our regression model involved finding the features best describing the influence of lexical and contextual predictability on word duration. One group of such factors consisted of word frequency and the collocational strength of the word with its left and right neighbour. Those factors were represented logarithmically, e.g.

$$word\ frequency = log\left(\frac{\text{number of occurrences}}{\text{total number of words in the corpus}}\right) \quad (2)$$

If the word happened to lie on the boundary of a prosodic phrase, i.e. if it was preceded or followed by a pause, its collocational strength with its left or right neighbour, respectively, was zero. Another group of factors was made up by characteristics describing the length of the word, its position in the phrase, some features of the word and its left collocate, and the length of the phrase. One of the features considered was the type (open vs. closed) of the last syllable of the left collocate, as vowel reduction has been observed on the final boundary of words ending in an open syllable [2]. The binary characteristic covers the monosyllabic feature. Word position is described within the phrase, while the phrase-final position is marked separately with a binary parameter. Table 3 presents the significance of all features depending on the part of speech (noun, adjective, conjunction and verb) of the (node) word analysed.

Conditionally the significance values (p) of the features in Table 3 can be divided into four groups: significant ($p<.05$), very significant ($p<.005$), highly significant ($p<.0005$) and not significant ($p>.05$). Obviously, across different parts of speech the influence of different factors displays a considerable variation. The strength of the right collocation, for example, is a highly significant feature for nouns, significant for adjectives, and not significant for verbs. Thus, even the significance patterns of content words (noun, adjective, verb) are rather variable, while no general content vs. function word opposition is revealed at all. On average the factors in the table have the least influence on the duration of conjunctions, as five features are insignificant for them. This may be explained by three reasons:

(1) conjunctions generally have a simple structure,

(2) many of them belong to uninflected words, and

(3) they often lie on the boundary of a prosodic phrase, where neighbour influence is less. Whether such high variation of the factors' influence across parts of speech is a regular feature or something specific to the sample studied requires additional research using much more voluminous speech material from many speakers.

Table 3: Factors of predictability and lexicality and their significance (p-values, NS = not significant p>.05) depending on part of speech.

Factor	Noun	Adjective	Conjunction	Verb
Frequency	<.05	NS	NS	<.05
Collocational strength with left collocate	NS	<.005	NS	<.05
Collocational strength with right collocate	<.0005	<.05	<.05	NS
Syllable type of left collocate	NS	<.05	<.05	NS
Length of word in phonemes	<.0005	NS	NS	<.0005
Position of word in phrase	NS	<.005	NS	NS
Last word in phrase	<.0005	<.005	<.05	<.0005
Monosyllabic word	NS	NS	<.0005	.05
Length of phrase in words	<.05	NS	NS	<.05

4. CONCLUSION

The study addressing a few very frequent Estonian words has revealed a slight correlation between collocational strength and word duration. The words co-occurring more frequently are pronounced shorter, while the left collocate has a slightly higher effect on the node word than the right one. Of lexical features, part of speech is significant for modelling word duration. It was revealed that proper names are pronounced the most slowly, whereas the most rapid pronunciation was found in ordinal numbers. The influence of factors of predictability and lexicality depends on part of speech, while the binary variable of content vs. function word is insufficient. By way of conclusion we state that although Estonian is an extremely word-bound language, the pronunciation of Estonian words in fluent speech is to a certain extent affected by some lexical categories as well as by the frequency and predictability of the words.

5. ACKNOWLEDGEMENTS

This work has been supported by the grant ETF7998 and project SF0050023s09.

6. REFERENCES

[1] Bell, A., Brenier, J., Gregory, M., Girand, C., Jurafsky, D. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1), 92-111.

[2] Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., Gildea, D. 2003. Effects of disfluencies, predictability and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113(2), 1001-1024.

[3] Campbell, N. 2000. Timing in speech: a multilevel process. In Horne, M. (ed.), *Prosody: Theory and Experiment.* Dordrecht/Boston/London: Kluwer Academic Publishers, 281-334.

[4] Piits, L., Mihkla, M., Nurk, T., Kiissel, I. 2007. Designing a speech corpus for Estonian unit selection synthesis. *Nodalida 2007 Proceedings: The 16th Nordic Conference of Computational Linguistics*, 367-371.

[5] Pluymaekers, M., Ernestus, M., Baayen, R.H. 2005. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62, 146-159.

[6] Sagisaka, Y. 2003. Modelling and perception of temporal characteristics in speech. *Proc. 15th ICPhS* Barcelona, 1-6.

[7] Vainio, M. 2001. *Artificial Neural Network Based Prosody Models for Finnish Text-to-speech Synthesis.* Helsinki: University of Helsinki.