Umeå University, Department of Philosophy and Linguistics
PHONUM 9 (2003), *149-152*
Available online at http://www.ling.umu.se/fonetik2003/

Evaluating rules for phonological reduction in Swedish

Per-Anders Jande

Centre for Speech Technology, Department of Speech, Music and Hearing, KTH

In this paper, pronunciation variation in Swedish due to speaking style and speech rate is discussed. A tentative rule system for segment-level reduction is currently being evaluated by letting subjects assess the naturalness of synthetic speech generated from canonical transcriptions and transcriptions reduced by the system, respectively. Results from experiments using short sentences with explicit control over the rules applied have shown that reduced forms are preferred at high speech rates (rates above the synthesis default rate), while there is no significant bias in preference between canonical and reduced forms at the synthesis default speech rate. Presently, longer and less controlled passages of synthetic speech are being evaluated using the same experiment set-up. Using text passages of varying degree of formality, this experiment allows for testing the effects of text formality on perceived naturalness.

1. Introduction and Background

Differences in speaking style influence many aspects of the linguistic message and the acoustic realisation. In informal, casual speech, the speaker often assumes that the listener has more background knowledge than in more formal conversations. Informal speech is thus often less explicit. The speaking style also affects the choice of words, the speech rate etc. The amount of phonological and phonetic detail in the realisation of an utterance is also affected, so that fast and informal speech is often more reduced than slow and formal speech. Having knowledge about how pronunciation varies in a language can be beneficial e.g. for improving the accuracy of automatic speech recognition (ASR) systems and for improving the naturalness of speech synthesis and for synthesising different speaking styles.

The ultimate goal of the research presented in this paper is to develop a non-application specific model of Swedish pronunciation variation due to speaking style and speech rate. For a general description of speaking style dependent pronunciation variation in Swedish, both phonological and phonetic level rules will have to be formulated. For this purpose, data-driven methods will be used on annotated spontaneous speech corpora. The focus will be on general aspects of pronunciation variation, rather than on variation due to dialect or individual factors.

Work on reduction rules for Swedish (focusing on sandhi rules) has been reported by Gårding (1974) and Eliasson (1986). Further, Bannert and Czigler (1999) have published a status report on studies of reduction patterns in consonant clusters of spontaneous speech. Bruce (1985; 1986) has built a reduction rule system mainly for vowel and syllable elisions, which uses stress patterns and the alternation between strong and weak syllables to predict elision. This research thus complements work on consonant cluster reduction. There are also studies of reduction in Swedish at the phonetic level. For example, Engstrand (1992) has studied the phonetic variation in natural Swedish discourse.

As a starting point for the present research, a tentative knowledge-based phonological level rule system building partly on the research mentioned above has been formulated. The purpose of this system is to a base from which a more elaborate rule system can be built.

The use of data-driven methods to explore pronunciation variation has been shown to be beneficial in many studies concerning other languages than Swedish, especially for expanding ASR lexica. Using speech corpora and data-driven methods facilitates finding non-trivial correlations. Data-driven methods will be used in the further development of the rule system.

2. A Tentative Rule System

Building partly on the work of Gårding (1974) and Eliasson (1986), a tentative rule system for reduction of Swedish words has been constructed. The input to the system are phonological lexicon transcriptions corresponding to canonical pronunciations. The tentative rule system thus only concerns phonological level reduction, which means that only reduction phenomena that changes the number of segments or the segment identities in a segment string are considered. Vowel length and word stress and accent are features that are provided in the input transcriptions and reduction involving these features was also included in the phonological level reduction rules. Some vowel and syllable elision rules similar to those described by Bruce (1985; 1986) were thus possible to include in the system.

The system includes rules for haplology, general forward assimilation, recursive retroflex assimilation, backward assimilation, /r/ elision, /h/ elision, reduction of common suffixes, reduction of common stems, vowel reduction and syllable elision.

A haplology rule deletes the first of two identical consonant clusters at each side of a compound boundary. The forward assimilation rules transfer phonological features or sets of features from a segment in a phonological string to the succeeding segment, so that the segments come articulatory closer. The recursive retroflex assimilation is a special case of forward assimilation. In central and northern Swedish dialects, the retroflex feature tends to spread recursively rightwards to consecutive adjacent dentals. Backward assimilation works in the same way as forward assimilation, but with a segment affecting its preceding neighbour instead of its succeeding neighbour. Assimilation can result in merges, so that the number of segments in the phonological string is reduced.

The /r/ elision and /h/ elision rules delete /r/ and /h/, respectively, in certain contexts. The special rules for common suffixes (e.g. noun inflections) and word stems handle reduction phenomena that are specific to certain units. The vowel reduction rules reduce unstressed vowels to schwa in certain contexts and shorten most long unstressed vowels. Syllable elision rules delete certain syllables. The vowel reduction and syllable elision rules use word-internal stress patterns as context.

3. Evaluation Approaches

There are many possible ways of evaluating the rule system. For example, the result of applying the rules to canonical transcriptions can be compared to transcriptions of actual speech to determine the predictive power of the system. This can be done for development purposes, to detect possible flaws in the system that need correction. For automatic speech recognition purposes, the system can be evaluated through generating alternative pronunciations and adding them to a ASR lexicon and test differences in recognition accuracy compared to a system with only canonical transcriptions in the lexicon. The system can also be evaluated from a synthesis perspective, testing the effects on intelligibility, conceptual load and perceived naturalness of applying the rules to input transcriptions before synthesis.

Testing perceived naturalness is easier than testing intelligibility or conceptual load and it is probable that a synthesis perceived as more natural is also easier to understand and less demanding to listen to. For this reason, initial evaluation of the rule system quality is done through letting subjects assess the naturalness of synthetic speech subjected to the reduction system.

4. Results

In a recently completed experiment (Jande, 2003), fifteen subjects listened to pairs of stimuli, where both stimuli were synthetic readings of the same sentence, one in canonical form and one in reduced form. Each pair was presented with three different speech rates (the synthesis default rate and two faster rates). The subjects were to mark the most natural sounding stimulus of each pair.

The results showed that the reduced forms of the sentences were perceived as more natural significantly more often at the higher speech rates. When the speech rate was low (synthesis default), there was no significant difference in perceived naturalness between the reduced and the canonical forms. The perceived naturalness was significantly dependent on speech rate, so that the preference for the reduced forms increased with increasing speech rate. Sentences with heavily reduced low frequency words, however, were perceived as less natural than their canonical counterparts also at the higher speech rate. This suggests that word predictability should be taken into account at rule application.

5. Work in Progress

The experiment reported in Jande (2003) used stimuli that were closely controlled, so that a wide range of the rules in the system were known to take effect. The sentences also differed only in one word, so that the causes of preference biases for specific sentences deviating from the general pattern could be more easily targeted. However, it is also interesting to investigate how the reduction rule system affects the perceived naturalness of synthetic speech in general.

For the purpose of such a general system evaluation, the same experiment set-up as in Jande (2003) is presently being used to evaluate the perceived naturalness of larger chunks of speech. Ten passages with an average of 60 words per passage were synthesised in canonical and reduced form at three different speech rates using an experimental version of the Infovox 330 diphone Swedish male MBROLA voice implemented as a plug-in to the WaveSurfer speech tool (cf. Beskow & Sjölander, 2000).

The passages used were excerpts from texts used in an experiment by Carlson et. al (1992) and there was thus no specific control over which rules would apply. The passages used differ in degree of complexity and formality. Four of the texts are translations of English texts used for comprehension test for college-level readers and six texts are translations of English texts written for fourth-grade readers. The texts directed to college students are more formal, while the texts directed to fourth-graders are less formal and have some elements of direct speech (e.g. exclamations and rhetorical questions). The formal texts have longer sentences and longer, more abstract words than do the less formal texts. These differences are clearly reflected in the passages used in the experiment. For example, the informal passages contain in average 9.85 words per sentence while the formal passages contain 17.5 words per sentence.

The tentative rule system is constructed primarily for conversational speech, not for “read speech”. This is hypothesised to be reflected in the results of the study. The less formal passages are syntactically more similar to conversational speech. The hypothesis is thus that the less formal passages will have preference biases more towards the reduced speech than will the more formal passages.

Preliminary results show patterns similar to those reported in Jande (2003). However, it is too early to say whether the results can support the hypothesis about the degree of formality affecting the preference bias.

6. Acknowledgements

The research reported in this paper was carried out at the Centre for Speech Technology, a competence centre at KTH (Royal Institute of Technology), Stockholm.

7. References

Bannert, R. & Czigler, P. E. (1999) Variations in consonant clusters in standard Swedish, *Phonum 7, Reports in Phonetics*, Umeå: Umeå University.

Beskow, J. & Sjölander, K. (2000) WaveSurfer - a public domain speech tool. In *Proceedings of ICSLP 2000*, pp. 464-467

Bruce, G. (1985) Fonologiska regler för elliptiskt tal (Phonological rules for elliptic speech). In *Svenskans beskrivning 15* (S. Allén, editor), pp. 149-158. Gothenburg: Gothenburg University.

Bruce, G. (1986) Elliptical Phonology. In *Papers from the Ninth Scandinavian Conference on Linguistics* (Ö. Dahl, editor), pp. 86-95. Stockholm: Stockholm University

Carlson, R., Granström, B., Neovius, L. & Nord, L. (1992) The “listening speed” paradigm for synthesis evaluation. In *Fonetik’92, Sixth Swedish Phonetics Conference, Chalmers Technical Report No 10*, pp. 63-66. Gothenburg: Department of Information Theory, Chalmers University of Technology.

Eliasson, S. (1986) Sandhi in peninsular Scandinavian. In *Sandhi phenomena in the languages of Europe* (H. Andersen, editor), pp. 271-300. Berlin: Mouton de Gruyter.

Engstrand, O. (1992) Systematicity of phonetic variation in natural discourse, *Speech Communication*, 11, 337-346.

Gårding, E. (1974) Sandhiregler för svenska konsonanter (Sandhi rules for Swedish consonants). In *Svenskans beskrivning 8* (C. Platzack, editor), pp. 97-106. Lund: Department of Nordic Languages, Lund University.

Jande, P. A. (2003) Phonological reduction in Swedish. In *Proceedings of ICPhS 2003* (forthcoming)