Multimodality in language and speech systems. Dordrecht: Kluwer Academic Publishers, 173-207.
Ward, N., 2006. Methods for discovering prosodic cues to turn-taking. *Proceedings Speech Prosody 2006*, Dresden.
Ward, N. & W. Tsukahara, 2000. Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics 32*, 1177-1207.
Wesseling, W. & R.J.J.H. van Son, 2005. Early preparation of experimentally elicited minimal responses. *Proceedings Sixth SIGdial Workshop on Discourse and Dialogue*, ISCA, Lisbon, 11-18.

Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics
Working Papers 52 (2006), 57–60

SMTC – A Swedish Map Task Corpus

Pétur Helgason
Department of Linguistics and Philology, Uppsala University
petur.helgason@lingfil.uu.se

Abstract

A small database of high quality recordings of 4 speakers of Central Standard Swedish is being made available to the speech research community under the heading Swedish Map Task Corpus *(SMTC). The speech is unscripted and consists mostly of conversations elicited through map tasks. In total, the database contains approximately 50 minutes of word-labelled conversations, comprising nearly 8000 words. The material was recorded at the Stockholm University Phonetics Lab. This paper describes the recording method, the data elicitation procedures and the speakers recruited for the recordings. The data will be made available on-line to researchers who put in a request with the author (cf. section 7 below).*

1 Introduction

The data being made available under the heading *Swedish Map Task Corpus* (SMTC) were originally recorded as part of the author's doctoral dissertation project (Helgason, 2002). The data have already proved useful for several other research projects, e.g. Megyesi (2002), Megyesi & Gustafson-Čapková (2002) and Edlund & Heldner (2005). As it seems likely that future projects shall want to make use of the data, and the data are not described in much detail elsewhere, an account of the recording procedure and elicitation method are called for. At the same time, the data shall be made available for download for researchers.

2 Recording set-up

The data were recorded in the anechoic room at the Stockholm University Phonetics Lab. The subjects were placed facing away from one another at opposite corners of the room (see Figure 1). The "head-to-head" distance between the subjects was approximately two meters. The reason for this placement of the subjects was partly to minimize cross-channel interference, and partly to prevent them from consulting one another's maps (see the following section). The recording set-up was therefore in accordance with the nature of the data elicitation method.

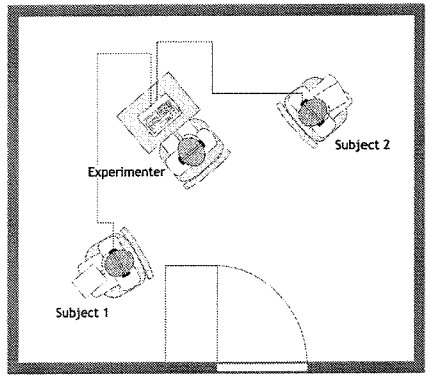

Figure 1. The placement of subjects and experimenter during the recording.

The data were recorded using a Technics SV 260 A DAT recorder and two Sennheiser MKE2 microphones. Each microphone was mounted on a headset and placed in such a way that it extended approximately 2.5 cm out and to the side of the corner of the

subject's mouth. The recording device and an experimenter were placed in between the subjects, within the anechoic room.

The subjects were recorded on separate channels. This was done in order to avoid an overlap between the subjects when they were speaking simultaneously. The absorption of sound energy in the anechoic room proved to be quite effective. The difference in average RMS between speakers on a channel was approximately 40 dB. Thus, for example, the average RMS for the intended right channel speaker was, on average, 40 dB higher than for the interfering (left-channel) speaker. This means that at normal listening levels (and provided the intended speaker is silent), the interfering speaker can be detected only as a faint background murmur.

3 Data elicitation – the map tasks

Most of the data were elicited by having the subjects perform map tasks. Map tasks have previously been used successfully for eliciting unscripted spoken data, perhaps most notably in the HCRC Map Task Corpus (Anderson et al., 1991). A map task involves two participants, an instruction giver and an instruction follower (henceforth Giver and Follower). For each map task, the experimenter prepares two maps with a set of landmarks (symbols or drawings), and to a large extent the landmarks on the two maps are the same. However, some differences in landmarks are incorporated by design, so that the maps are not quite identical. The Giver's map has a predetermined route drawn on it, the Follower's map does not. Their task is to cooperate through dialogue, so that the route on the giver's map is reproduced on the follower's map. The Giver and Follower are not allowed to consult one another's maps. In the SMTC recordings, the subjects were told at the beginning of the task that the maps differed, but it was left up to them to discover the ways in which they differed.

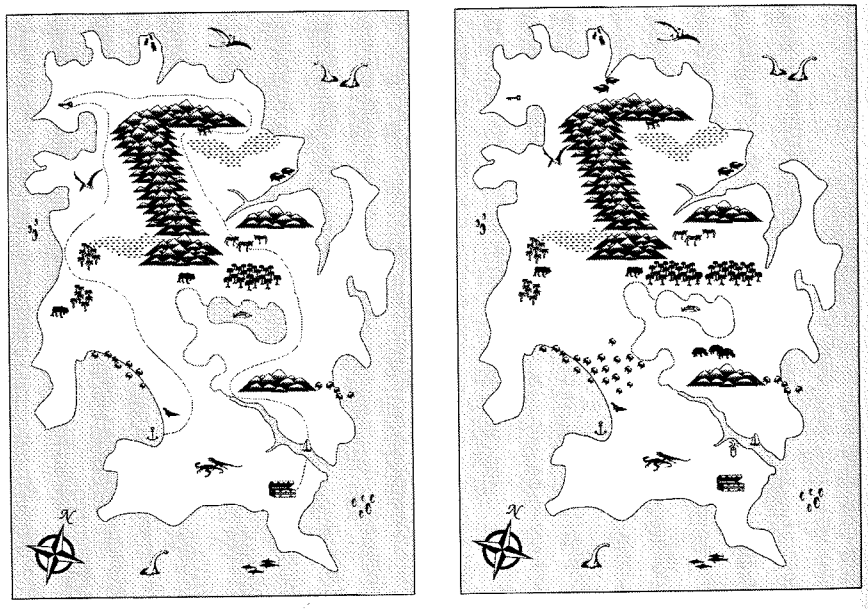

Figure 2. One of the "treasure hunt" map pairs used for data elicitation. On the left is a Giver's map, and on the right is a Follower's map. The path on the Giver's map is reproduced in grey here, but when the subjects performed the tasks it was marked in green.

For the SMTC recordings, four map tasks were prepared, each consisting of a Giver and Follower map pair. An example of a map task is given in Figure 2. All the maps had a set of basic common features. They all depicted the same basic island shape, the contour of which had several easily recognizable bays and peninsulas. The island also had mountains and hills, as well as a lake and a river. Finally, each map had a simple compass rose in the bottom left corner.

Two of the map tasks had a "treasure hunt" theme. The landmarks on these maps included an anchor (the starting point), a key (an intermediate goal), and a treasure chest (the final goal). However, most of the symbols depicted animals (some of which were prehistoric) and vegetation.

The remaining two map tasks had a "tourist" theme. On these maps the landmarks consisted entirely of various symbols typically used in tourist brochures and maps. In order to familiarize the subjects with these symbols, they were asked to go through a list of such symbols with a view to deciding how to refer to them if they occurred in a map task. This interaction was recorded, and is included in the SMTC database under the heading "Symbol task". This symbol list is reproduced here in Figure 3. The subjects' goal in the tourist maps was to trace a predetermined route around the island from an airport and back to the same airport.

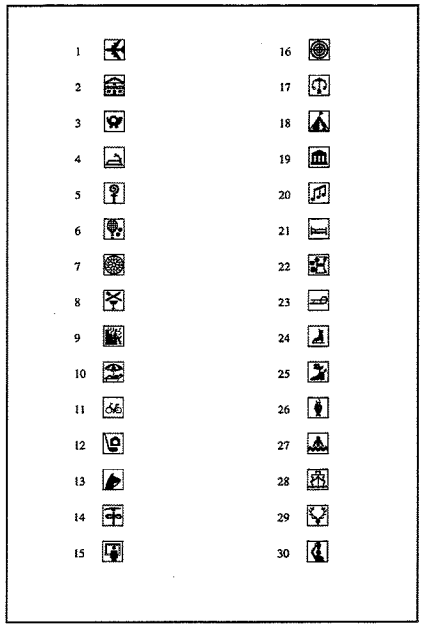

Figure 3. The symbol list used to familiarize the subjects with the symbols on the "tourist theme" maps.

4 The subjects

The subjects, one male and three females, were recruited from the staff at the Stockholm University Linguistics Department. They are referred to as FK, FT, FS (all female) and MP (male). FK and MP were in their thirties and FT and FS in their forties. All speakers were of normal hearing. As regards dialect, all speakers identified themselves as speakers of Central Standard Swedish and had lived for most or all of their lives in or around Stockholm. They were paid a moderate fee for their participation.

The subjects were arranged in pairs of two, FS and MP as one pair and FK and FT as another. Each pair began the session by navigating through a "treasure hunt" map task. This was followed by a discussion of the symbol list. The pair then continued with a "tourist map" followed by another "treasure map", and finally, if time allowed, one more "tourist map".

5 The extent of the database

The data from both subject pairs comprise a total of approximately 50 minutes of conversation. (There exist additional map-task recordings of these as well as other subjects which await word-labelling, but these are not included in the present database). This represents a total of 35 minutes of uninterrupted speech from the four subjects. (What is referred to here as

uninterrupted speech is the total speaking time for a subject, excluding any and all pauses.) For FT, approximately 4.3 minutes of uninterrupted speech are available, comprising a total of 870 words; for FK 9.5 minutes comprising 2045 words; for MP 10.8 minutes comprising 2554 words; and for FS 10.3 minutes comprising 2401 words.

6 Some remarks on the transliteration provided with the recordings

The data are provided with a word-level transliteration (word labelling). The transliteration was performed by the author, a non-native (albeit competent) speaker of Swedish. Researchers that wish to make use of the data may make use of this transliteration, possibly using it as the basis for searches or subject it to automatic text processing. Therefore, the rationale behind the transliteration conventions will be outlined here.

The aim of the transliteration was to facilitate lexical look-ups rather than to indicate or reflect the segmental content. For instance, the function word *det* is always indicated simply as "det" in the transliteration, without regard for any variability in its production (e.g. [de:t], [de:], [dɛ], [ɾe:] or [d̪e]). This approach was also applied in the labelling of minimal responses and lexical fillers. For example, lexical fillers of the "eh" or "er" type are indicated with a semicolon ; in the transliteration, irrespective of their segmental content (schwa-like, [e]-like, [œ]-like, creaky, nasalized, etc.)

A prominent feature of the transliteration is that contiguous pieces of speech (i.e. stretches of speech which contain no silence pauses) are demarcated at the onset and offset with a period (full stop) symbol. Thus the transliteration does not attempt to reflect the syntactic structure of an utterance, but instead only the presence of silence pauses. Note, also, that the transliteration provides no evaluation of or amendment to the grammaticality of an utterance.

7 Format and availability

The sound files are 16-bit stereo (with one speaker on each channel) and have a sampling rate of 16 kHz. The files are provided in the uncompressed Wave PCM format (i.e. *.wav). The word label files are provided as text files in WaveSurfer format. The data are made available as is, with no guarantee of groundbreaking research results. To obtain the data, please e-mail a request to the author to obtain a web address from which to download the data.

Acknowledgements

The author would like to thank Sven Björsten, Peter Branderud and Hassan Djamshidpey for their assistance with the recording of the data.

References

Anderson, A.H., M. Bader, E.G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson & R. Weinert, 1991. The HCRH Map Task Corpus. *Language and Speech 34*(4), 351–366.

Edlund, J. & M. Heldner, 2005. Exploring Prosody in Interaction Control. *Phonetica 62*(2–4), 215–226.

Helgason, P., 2002. *Preaspiration in the Nordic Languages: Synchronic and Diachronic Aspects*. Ph.D. Thesis, Stockholm University.

Megyesi, B., 2002. *Data-Driven Syntactic Analysis – Methods and Applications for Swedish*. Ph.D. Thesis, Department of Speech, Music and Hearing, KTH, Stockholm.

Megyesi, B. & S. Gustafson-Čapková, 2002. Production and Perception of Pauses and their Linguistic Context in Read and Spontaneous Speech in Swedish. *Proceedings of the 7th ICSLP*, Denver.

Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics
Working Papers 52 (2006), 61–64

The Relative Contributions of Intonation and Duration to Degree of Foreign Accent in Norwegian as a Second Language

Snefrid Holm
Department of Language and Communication Studies,
Norwegian University of Science and Technology (NTNU)
snefrid.holm@hf.ntnu.no

Abstract

This study investigates the relative contributions of global intonation and global segment durations to degree of foreign accent in Norwegian. Speakers of Norwegian as a second language (N2) from different native languages (L1s) plus one native Norwegian (N1) speaker are recorded reading the same sentence. The N2 utterances' global intonation and global segment durations are manipulated to match the N1 pronunciation. In this way every N2 speaker provides four utterance versions: the original, a duration corrected version, an intonation corrected version and a version with both features corrected. N1 listeners judge the degree of foreign accent between each speaker's four utterance versions. The results show that a) the combined correction of both features reduces the degree of foreign accent for all speakers, b) each correction by itself reduces the degree of foreign accent for all but two of the investigated L1 groups and c) some L1 groups benefit more from intonation correction whereas others benefit more from duration correction.

1 Introduction

When learning a second language after early childhood the resulting speech will normally be foreign accented (e.g. Flege, Munro & Mackay, 1995). The phenomenon of foreign accent is complex and comprises issues regarding the nature of the foreign accent itself as well as the foreign accent's various effects on listeners, for instance regarding social acceptance or the ability to make oneself understood.

A foreign accent may not in itself hinder communication. Although degree of foreign accent is often confounded with degree of intelligibility, a growing body of evidence supports the view that even heavily accented speech may sometimes be perfectly intelligible (Derwing & Munro, 1997; Munro & Derwing, 1995). The relationship between a deviating pronunciation on the one hand and its effect on listener dimensions like intelligibility or perceived degree of foreign accent on the other hand is not clear. There is however a general belief that prosodic deviations are more important than segmental ones, at least for intelligibility, although there are rather few studies to support this view (Munro & Derwing, 2005).

This study aims to establish which of the two pronunciation features global intonation and global segment durations contributes most to perceived degree of foreign accent in Norwegian as spoken by second language learners.

The present paper reports on a study which is part of a larger work where the next step will be to investigate the effect of the same two pronunciation features upon intelligibility. In this