

# A Danish phonetically annotated spontaneous speech corpus (DanPASS)<sup>☆</sup>

Nina Grønnum

*Linguistics Laboratory, Department of Scandinavian Studies and Linguistics, University of Copenhagen, 120 Njalsgade, DK-2300 Copenhagen, Denmark*

Received 30 August 2006; received in revised form 23 September 2008; accepted 4 November 2008

## Abstract

A corpus is described consisting of non-scripted monologues and dialogues, recorded by 27 speakers, comprising a total of 73,227 running words, corresponding to 9 h and 46 min of speech. The monologues were recorded as one-way communication with an unseen partner where the speaker performed three different tasks: (s)he described a network consisting of various geometrical shapes in various colours, (s)he guided the listener through four different routes in a virtual city map, and (s)he instructed the listener how to build a house from its individual pieces. The dialogues are replicas of the HCRC map tasks. Annotation is performed in Praat. The sound files are segmented into prosodic phrases, words, and syllables. The files are supplied, in separate interval tiers, with an orthographical representation, detailed part-of-speech tags, simplified part-of-speech tags, a phonemic notation, a semi-narrow phonetic notation, a symbolic representation of the pitch relation between each stressed and post-tonic syllable, and a symbolic representation of the phrasal intonation.

© 2008 Elsevier B.V. All rights reserved.

*Keywords:* Monologue; Dialogue; Spontaneous speech; Corpus; Phonetic notation; Prosodic labeling

## 1. Introduction

Most of our insight into the phonetics of spoken Danish to date is based on carefully manipulated, scripted material read aloud in a recording studio in the laboratory. This is not as strange as it may sound to non-phoneticians. First of all, even the largest non-scripted speech corpora may fail to exhibit a sufficient number of instances of the phenomenon to be investigated – in the proper context. Secondly, many phonetic phenomena are best studied when the variable under investigation can be carefully controlled and isolated from other – potentially interacting – phenomena. Thus, for example, the study of tone necessitates control over voicing and aspiration in consonants in the syllable

onset and over vowel quality/height, and any study of duration calls for control over stress and segmental context. Results obtained from manipulated read materials may serve – at a later stage – as a reference for data obtained from non-scripted speech. In brief, scripted materials read aloud in the laboratory may lack spontaneity but they can be made to meet legitimate, specific phonetic research requirements. However, there is a large number of interesting questions about connected speech that cannot be exhaustively answered from samples of scripted speech. This is especially true of reduction phenomena and of prosody, particularly prosody and its interaction with syntax and pragmatics.

Non-scripted speech may be obtained in various ways, each with its own advantages and disadvantages. It may be truly spontaneous and recorded in the speaker's natural environment, i.e. the experimenter exerts no control over what the speaker talks about or how, and the speaker avoids the slightly intimidating recording studio environment. This will presumably ensure a maximum of naturalness of speech.

<sup>☆</sup> The corpus was presented at the 5th International Conference on Language Resources and Evaluation, Genova 24–24 May 2006. A shorter version of this paper is included in the conference CD-ROM – see Grønnum (2006).

*E-mail address:* [ninag@hum.ku.dk](mailto:ninag@hum.ku.dk)

However, although eliciting speech in a recording studio may compromise naturalness somewhat, it has distinct advantages over spontaneous speech recorded in the field. Thus, a studio recording generally yields a better acoustic signal, essential for a number of phonetic analyses. Particularly, setting speakers specific tasks, i.e. specific subjects to talk about, as in this corpus, will facilitate comparisons and generalizations across speakers. Furthermore, since the speakers had to name specific landmarks in the maps, a direct comparison is made possible between the reduced forms of the non-scripted speech and the distinct forms produced in the subsequent reading aloud of the landmark names. A distinct advantage of the chosen procedure is also that corpora similar to this one already exist for other languages, opening the road to cross-language comparative studies – see, for example, Swerts (1994), Swerts and Collier (1992), Fletcher et al. (2002), Helgason (2006) and Horiuchi et al. (1999).

The intention was to supply a corpus for acoustic and perceptual phonetic investigations. That is, the primary goal is not syntactic, pragmatic, socio-linguistic, psychological, or any other specific aspect of spoken language one might wish to investigate. There are therefore a considerable number of discourse variables that have not been taken into account in the choice of elicitation material. Nevertheless, the corpus may serve as a basis for a number of linguistic and/or speech technological investigations. An obvious use is as training material for automatic segmentation and annotation, and it has in fact been used as such in the preliminary stages of an investigation of acoustic and perceptual building blocks in spontaneously spoken Danish – see Dau and Christiansen (2007).

## 2. The corpus<sup>1</sup>

### 2.1. Monologues

The monologues were recorded in 1996 and represent various types of instructions. The speaker was seated alone in the professional recording studio of the department and could communicate with the experimenter (the author) only via microphone and headphone. Once the subject had been instructed in the specific task, (s)he could no longer address the author with questions or comments. In other words, the monologues were recorded in one-way communication with an unseen partner who offered no feedback, whether it be in the form of questions or confirmation. Speakers were recorded with professional equipment (Sennheiser Microphone ME64, Revox A700, Agfa PEM368 tape). The analog recordings were later digitized and transferred to CD-ROMs at a sampling frequency of 48 kHz.

Each speaker performed three tasks:

- (S)he described a network consisting of various geometrical shapes in various colours – see Appendix A. It is an elaboration of Swerts and Collier (1992) network. It was specifically intended to reveal whether or not speakers look ahead and signal prosodically an upcoming utterance boundary prior to its actual occurrence. Since the colours cannot be discerned in grey-scale, English colour terms have been supplied.
- (S)he guided the author through four different routes in a virtual city map, *Slotsby* – see Appendix B, inspired by Swerts (1994). Again, English colour terms have been added to the map for the present purpose.
- Given a model of a house as well as its individual building blocks – see Appendix C – (s)he told the author – who had only the individual pieces – how to assemble them. This house is an almost exact copy of Terken's edifice – see Terken (1984). English colour terms were subsequently supplied.

#### 2.1.1. Speakers

There were 18 speakers, 13 men and 5 women, all of them students or colleagues in the (former) Department of General and Applied Linguistics, all except one originating in the greater Copenhagen area. At the time of recording they were aged 68, 46, 41, 39, 35, 34, 33, 31, 30, 28, 26, 24, 23 (2), 22 (2), 21, and 20 years, respectively, i.e. 3 were older than 40 years and 15 were younger. None of them had any known speech or language deficits.

#### 2.2. Dialogues

The dialogues were recorded in the summer of 2004. They are replicas of the Human Communication Research Centre's Map Tasks – see Anderson et al. (1991), Brown et al. (1984) and <http://www.hrcr.ed.ac.uk/maptask/>.

The exercise involved the co-operation of two participants. They were seated in separate locations, one in the department's recording studio, the other in a recording facility established for the purpose in the main control room with curtains of very heavy material surrounding the speaker. The speakers communicated via headsets.

A laboratory set-up like this is hardly the most natural environment for communication, but it turned out to be necessary in order to obtain recordings of sufficiently good quality for subsequent acoustic analysis: seated in the same room, across from each other with eye-contact, speaker A could invariably be heard over speaker B's microphone, and vice-versa, whereas clean acoustic signals were obtained when the speakers were separated, with no appreciable difference in quality from the studio proper and the ad hoc studio established in the control room. Given the setting, i.e. the lack of visual and direct auditory contact, the participants would presumably be more comfortable if they were not also required to communicate with a

<sup>1</sup> For complete and detailed information about speakers, processing and annotation conventions see the website, <http://www.danpass.dk>.

Download English Version:

<https://daneshyari.com/en/article/566154>

Download Persian Version:

<https://daneshyari.com/article/566154>

[Daneshyari.com](https://daneshyari.com)