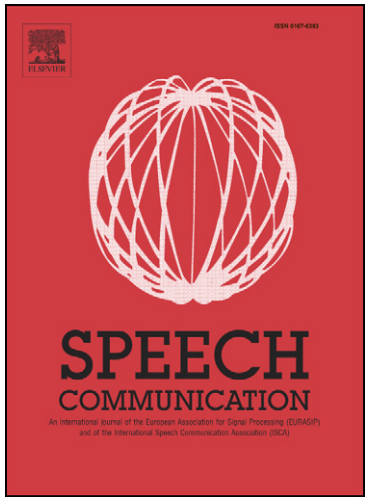
A Danish Phonetically Annotated Spontaneous Speech Corpus (DanPASS)

Nina Grønnum



A Danish Phonetically Annotated Spontaneous Speech Corpus (DanPASS)

Nina Grønnum

PII: S0167-6393(08)00174-X
DOI: 10.1016/j.specom.2008.11.002
Reference: SPECOM 1760

To appear in: *Speech Communication*

Received Date: 30 August 2006
Revised Date: 23 September 2008
Accepted Date: 4 November 2008

Please cite this article as: Grønnum, N., A Danish Phonetically Annotated Spontaneous Speech Corpus (DanPASS), *Speech Communication* (2008), doi: 10.1016/j.specom.2008.11.002

A Danish Phonetically Annotated Spontaneous Speech Corpus (DanPASS)[1]

Nina Grønnum
Linguistics Laboratory, Department of Scandinavian Studies and Linguistics, University of Copenhagen
120 Njalsgade, DK-2300 Copenhagen, Denmark
E-mail: ninag @ hum.ku.dk

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 hours and 46 minutes 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.

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

[1] 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).

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. 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), 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[2]

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's (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 (1984) edifice. 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] For complete and detailed information about speakers, processing and annotation conventions see the website, http://www.danpass.dk.

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.hcrc.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 stranger. Accordingly, the two members of a pair knew each other well. They were recorded via professional headset microphones (Voice Technologies VT700), directly onto CD-ROMs (HHB Professional Compact Disc Recorder CDR-850) to separate channels in a stereo recording.

Each participant had a map. One, the giver, had a route on his or her map; the other, the follower, did not. Their goal was to collaborate so as to reproduce the giver's route on the follower's map. The maps were not exactly identical: landmarks were missing on one or the other map, a landmark might appear twice on one map but in only one location on the other, and a given landmark might have slightly different names on the two maps – see the example in appendix D. This, of course, is what gave rise to a true negotiation, with questions and answers, backtracks and repairs. Participants were informed explicitly in writing about these irregularities prior to the recording. It was left to them, however, to discover how and where the maps or the designations differed, and to supply the missing items and correct names on their respective maps. Each pair of speakers completed four different sets of maps.

It was our distinct impression, during the subsequent processing of the recordings, that the speakers had been comfortable with the task and the experimental setting. They produced fluent speech for monologues as well as dialogues and were not in any obvious way influenced by the non-naturalness

of the circumstances.

2.2.1. Speakers

22 speakers participated, 13 of whom also recorded the monologues in 1996. They were all from the greater Copenhagen area, drawn from the pool of (former) students and colleagues. There were 13 men and 9 women, aged 76, 62, 59, 58, 54, 49, 47 (2), 42, 41, 38, 36, 31, 30(4), 28 (2), 27(2), 22, i.e. 10 were over and 12 were under 40 years of age.

2.3. Word Lists

After completion of the map sessions, subjects were asked to read a word list containing all the feature names from the maps they had encountered. Each name appeared twice, in random order, and subjects were asked to read the list in a distinct speech mode. The lists provide citation forms for comparison with the less distinct dialogue forms. Landmarks and names in the original English maps were designed with specific phonological phenomena and processes in mind. The translation into Danish was constrained by the nature of the landmarks, with only moderate influence over phonological structure.

2.4. Video Recording

In the studio proper a video-recorder was mounted. The camera was placed as close as possible, and as nearly perpendicular as possible, to the frontal plane of the speaker's face without impeding his/her view of the map. The videos were intended as analytical material for anyone who should want to attempt to accompany synthetic Danish speech with a model talking face.

Each speaker had to serve as giver as well as follower, in alternation. Each speaker also had to be video-recorded in both roles. Accordingly, after two map sessions, with speaker A being giver and follower, respectively, the speakers changed places in order for speaker B to be video-recorded as well. Thus, each pair of speakers had to run through four different sets of maps. A complete recording session lasted between 30 and 40 minutes.

2.5. Statistics

There are 9 hours and 46 minutes of speech altogether, 2 h 51 m in the monologues and 6 h 55 m in the dialogues. There are 2121 different word forms in the corpus as a whole, 1075 in the monologues and 1593 in the dialogues. There are 21 170 running words in the monologues and 52 057 in the dialogues,

i.e. a grand total of 73 227 running words in the corpus.

[Figure 1 about here]

3. Processing

The speech signals were processed in Praat – see Boersma (2001) and Boersma and Weenink (2006). Fig. 1 presents a screen shot of a section of one of the city map monologues. There are 10 separate interval tiers for (1) the orthographic transcript, (2) detailed part-of-speech (POS) tags, (3) simplified POS-tags, (4) a phonological notation, (5) a semi-narrow phonetic notation within the word domains, (6) the same semi-narrow phonetic notation within the syllable domains, (7) a symbolic representation of the pitch relation between each stressed syllable and its first post-tonic syllable, (8) a symbolic representation of the phrasal intonation contour. Tier 9 is for comments. In a project headed by Patrizia Paggio at the Centre for Language Technology, University of Copenhagen, the information structure of the monologues was analysed and topic and focus tags added to the orthography in a separate tier at the bottom (10).

There is a search engine attached to the corpus. It will perform searches in any of the ten tiers and also allows for combined searches in different tiers as long as the temporal domains are of equal magnitude. Thus, tier 5 was introduced in order to permit combined searches in the phonological and phonetic representations.

3.1. Segmentation

Marking boundaries in the sound files at the level of individual phonetic segments would have been ideal but proved impossible, given constraints of time and money. The smallest delimited temporal domains are therefore syllables. The next larger temporal domains are words, and the largest delimited domains are prosodic phrases. Prosodic phrase boundaries were always made to coincide with word boundaries, and word boundaries always coincided with syllable boundaries. Boundaries were always located in the nearest zero-crossing in the signal, a procedure which ensured that no spurious phantom clicks were heard in the signal when we replayed individual syllable or word domains.

Segmentation was mostly straightforward and uncontroversial. However, due to the specific phonetic and phonological properties of Danish there are polysyllabic sequences which cannot be delimited. Thus, for example, there is no non-arbitrary way to determine syllable boundaries in words

like *gade, løve, sagde, køre, dreje* "street, lion, said, drive, turn." Phonologically they are /gaːdə løːvə saːgə køːrə drajə/, phonetically they are [ˈg̊æːð̩ ˈløːʊ ˈsæːæ ˈg̊ʰøːɐ ˈd̥ʁɑɪ].[3] The initial onset consonant(s) is/are succeeded by one long vocalic sound, whether it be stationary or non-stationary. The words are bi-syllabic, partly by virtue of their total duration but principally by their fundamental frequency pattern. Similarly, boundaries between vocalic sounds on either side of a word boundary are of course impossible to determine.

Prosodic phrase boundaries presented a problem of a fundamental nature when they did not coincide with utterance boundaries. There are no unambiguous, consistent, and objective acoustic cues to prosodic phrase boundaries in Danish. Of necessity, then, the guiding principle had to be purely auditory: the delimitation should result in domains which were perceived to be internally coherent and uniform with respect to their rhythm and their intonation. Conversely, the boundaries between prosodic phrases should be perceived as ruptures in the rhythm and intonation contours. Note specifically that prosodic phrase boundaries are not necesssarily accompanied by pauses nor do pauses occur exclusively at prosodic phrase boundaries – they may as well occur internally in prosodic phrases. Likewise, prosodic phrase boundaries may coincide with syntatic boundaries but they do not invariably do so nor does a syntactic boundary invariably introduce a prosodic boundary.

3.2. Annotation

Monologues and dialogues were transcribed orthographically in standard orthography without punctuation and using capital letters for proper names only. Empty pauses, filled pauses, and articulatory hesitation were indicated. Subsequently, the transcript was fed into the slots in tier 1. The orthographical representation is supplemented with stress marks – in the shape of a comma directly before the vowel letter representing the vowel of the stressed syllable – intended for researchers who are interested only in the distribution of stress across the texts, regardless of the pronunciation.

The POS-tagging in tiers 2 and 3 is automated. The tagger, developed by Peter Juel Henrichsen, Department of Computational Linguistics at Copenhagen Business School, was trained on written language – see Henrichsen (2002). At the outset there was no way to predict how well the tagger would perform on non-scripted speech. On the whole, the tagger turned out to be efficient and reliable, as revealed in the subsequent proof-reading of the entire corpus. But there were mistakes, some of them

[3] [ð̩] denotes a syllabic approximant, i.e. it is not a consonantal sound.

random and some of them systematic. For instance, *ja* "yes" and *nej* "no" were almost consistently labelled <noun>, whereas *næh* (an informal form of *nej*) was correctly labelled <interjection>. A somewhat more troublesome shortcoming was the fact that the category <article> was simply lacking from the tagger's inventory. Indefinite articles were labelled personal pronouns and definite articles were labelled demonstrative pronouns. Corrections had to be made manually.

The phonological notation in tier 4 was fed automatically into the word domains from a "phoneme dictionary," i.e. from a list of the 2121 different word forms in the corpus and their corresponding, manually supplied, phonological representations. The representation is fairly abstract where the segments are concerned, in accordance with the phonological analysis of Danish in Grønnum (2005), but stress marks are added to polysyllables and stød[4] is designated as well, although both stress and stød are to a very large extent predictable from the segmental and morphological structure and thus – strictly speaking – phonologically redundant. Adding stress and stød, however, will presumably facilitate certain search procedures at a later stage.

The phonetic notation in tiers 5 and 6 is broad where the stop consonants are concerned and semi-narrow, with a fairly liberal use of the relevant diacritics, where all other consonants as well as the vowels are concerned. Thus, [p t k] are convenient simplifications, broad notations, for [b̥ʰ d̥ˢ g̊ʰ] (unvoiced weak aspirated/affricated stops), and [b d g] are convenient simplifications, broad notations, for [b̥ d̥ g̊] (unvoiced weak unaspirated/unaffricated stops).

The symbolic representation in tier 7 of the pitch relation between stressed and first post-tonic syllable is graded in seven steps: the post-tonic is perceived to be either much higher (H/), higher (H), a little higher (h), equal to (=), a little lower (l), lower (L), or much lower (L\) than the stressed syllable. The interval is specified to such a relatively fine degree because in its magnitude lies a correlate to perceived prominence – see Grønnum (1990) and Jensen and Tøndering (2005).

Stressed syllables are labelled with a star (*) in tier 7. Among the 40 086 stressed syllables in the corpus some (2 467, i.e. about 6 %) were perceived (by one assistant and the author) to be more prominent than others. There are no unambiguous acoustic cues to extra prominence, but the author supplied post-hoc auditory characterizations of these syllables and reached a total of 12 different cues to extra prominence, often in combinations of two or three properties. (1) Prominence on

---

[4] There is no adequate English term for stød. It is a special kind of creaky voice characterizing certain syllable types under certain morphological conditions. See, for example, Grønnum and Basbøll, 2007.

succeeding stressed syllables may be reduced (red). (2) Greater loudness than in neighbouring stressed syllables may occasionally be involved (loud). (3) The rise to the post-tonic may be relative larger than in surrounding stress groups (h/). (4) Duration may be greater than in neighbouring stressed syllables (dur). (5) If the syllable onsets with a vowel it may be preceded by a glottal stop (?). (6) If the extra prominence is on the last stressed syllable in the phrase that syllable may be on a high pitch without sounding as a cue to continuation (hi%). (7) The pitch interval to the succeeding lower stressed syllable may be large (h-l). (8) The post-tonic syllable may be at a considerably lower pitch (l\). (9) The prominent syllable may be at a clearly higher pitch level than surrounding stressed syllables (hi), or (10) it may be at a clearly lower pitch level (lo). (11) There may be a clearly dynamic pitch movement within the stressed syllable (dyn). (12) The phrasal contour may not decline after the prominent syllable (>h). By far the most common characteristics of syllables perceived to have extra prominence are either greater duration (dur) or higher pitch level (hi). Syllables with extra prominence are labelled with an exclamation mark before the star in tier 7 (!*), and the auditory characterizations can be found in tier 9 – see fig. 1. Without being wholly arbitrary, the distinction between 'normal' and 'extra prominence' is not always unambiguous. We have perhaps been rather conservative when assigning 'extra prominence' to stressed syllables.

Intonation in the prosodic phrases in tier 8 is characterized, firstly, by the way the stressed syllables are pitch scaled throughout the phrase, i.e. by their mutual relationship, and, secondly, presumably also by the way the phrase onsets and offsets, i.e. by the pitch of the very first and very last syllable in the phrase, be it stressed or unstressed. The pitch of the stressed syllables and the syllables at the phrasal boundaries is represented on a broad scale of high (h), mid (m) and low (l). However, the means also exist to obtain finer gradations within a succession of stressed syllables in a given range – between high and mid, high and low, and mid and low. For instance, h_>_>_>_m designates a succession of five stressed syllables which descend gradually from high to mid.

Readers familiar with the ToBI convention for transcribing prosody, e.g. Silverman et al. (1992), should note that any similarity with our annotation is merely superficial. For the description of Danish intonation the phonological assumptions behind ToBI are inappropriate, and as a phonetic transcription system it is not sufficiently fine grained for our purpose – see Grønnum (1985, 1986, 1995). For a general critique of ToBI see Kohler (2005, 2006, 2007).

Note that, again for reasons to do with time and resources, the pitch relation between successive prosodic phrases is not represented. Given the flexibility of Praat, it can easily be added to the grid if

and when the need arises.

3.2.1. Annotation procedure

The segmental notation in tiers 5/6 and the prosodic labeling in tiers 7 and 8 were always done by two project assistants,[5] independently of each other and in parallel, in three stages: first the semi-narrow segmental notation, then the stress-and-pitch relation and finally the phrasal intonation. At each stage the two assistants met at regular intervals and compared their annotations, file for file, speaker for speaker. Disagreements between them were resolved in discussions with the author. Subsequently, the author proof-read the entire file. Over the years, three pairs of assistants were employed. They were all students of linguistics with a special orientation in phonetics. From the sessions with each pair it emerged that certain phenomena gave rise to more disagreements between assistants than others. This was particularly true of stress. Where the two assistants differed, the author served as arbiter. During the subsequent proof-reading, the annotation was occasionally modified. There are no objective measures of the validity of the annotations, but – given the overall rather good agreement between assistants, and given the repetitive procedure, i.e. the fact that we have listened to the recordings during multiple stages – the annotation may be considered a fairly adequate symbolic representation of the speakers' speech, adequate, that is, for phonetic investigations of non-scripted speech in Danish. However – and perhaps needless to say – phonetic notation, specifically of the rather narrow kind, and prosodic labeling are both impressionistic exercises and the true validity and adequacy of the corpus (and likewise its short-comings) will only become apparent when students and researchers use it for their various purposes.[6]

4. Preliminary analyses of the corpus

4.1. Prosodic phrases in the monologues

Tøndering (2008) performed an analysis of the prosodic phrases in the monologues. A phrase might contain only one stressed syllable, and, at the other extreme, there was a phrase with a total of 14 stressed syllables – see fig. 2. The bulk of the phrases contained either 2 or 3 or 4 stressed syllables, with an average of 3.4 stressed syllables per prosodic phrase.

---

[5] Except for extra prominence which involved only one assistant and the author.

[6] A number of students have used the corpus in their investigations – see the corpus website, www.danpass.dk. They have occasionally found mistakes or inconsistencies which have then been corrected.

[Figure 2 about here]

Tøndering also found that stressed syllables on mid level pitch were in the majority. Furthermore, there were more highs (than mids and lows) in the onset and more lows (than mids and highs) in the offset of a prosodic phrase, indicative of a general downdrifting trend. The pitch of onset stressed syllables was not – in phrases of four or more stressed syllables – correlated with phrase length, i.e. there were not significantly more high onsets in longer than in shorter phrases. When onsets thus were quasi-constant, differences in phrase length must entail differences in overall downdrift: longer phrases had less steep gradients than shorter phrases. This is in accordance with results for read speech – see Grønnum (1985, 1986, 1995).

The prevailing general downdrifting trend in the contours could be construed as a result of the fact that the monologues contained no questions.

[Figure 3 about here]

4.2. Question intonation in the dialogues

Intonation contour slopes in Danish read speech have been found to vary systematically with utterance type or modality – see Grønnum (1995). Declarative statements have the steepest gradients, wh-questions are slightly less steep, questions with word order inversion less steeply falling again, and so-called declarative questions have no gradient at all, i.e. their global contour is high and level – see fig. 3.

Grønnum and Tøndering (2007) analysed 300 questions and 51 statements from 24 dialogues to see whether speaker strategy carries over from scripted to non-scripted speech, i.e. whether a trade-off between lexicon and syntax vs. intonation contour slope could be found also in non-scripted speech. In that case the 51 statements should have the steepest gradients, the 47 wh-questions in the material should have slightly less steep gradients, the 114 questions with word order inversion should have even less declining slopes, and the 139 declarative questions should have level contours. Given the fact that intonation contour onsets do not vary systematically with either utterance type or utterance length, the contour gradients are adequately reflected in the contour offsets, i.e. declaratives should terminate at the lowest pitch level, wh-questions above them, questions with word order inversion higher again, and declarative questions on top, at approximately the same value as the contour onset. By and large, this

turned out to hold true, but among the 139 declarative questions there was a subgroup of 41 utterances which acoustically and perceptually were indistinguishable from the true declaratives, i.e. their slopes were coincident with the statements. Spliced out from their context and listened to in isolation they also sounded like perfectly ordinary statements. This is curious since one would expect that an utterance which has no overt lexical or syntactic markers of its interrogative function would have to have a prosodic cue, a non-statement-like intonation contour, in order to be perceived as a question. But apparently, under the proper circumstances, this is not a prerogative. What those circumstances are is not clear: we found no systematic contextual differences between these 41 declarative questions and the prosodically more conventional 98 declarative questions. We do believe, however, that this option – to ask a question without in any way sounding like it – has to do with the specific task and with the fact that the partners in each dialogue knew each other rather well.

4.3. Phrasal intonation in subordinate clauses in the monologues

Dyrby et al. (2005) investigated subordinate clauses in 8 house-building monologues. They found, among other things, that a typical subordinate clause was in utterance final position and had a falling intonation contour. If it was syntactically integrated with the matrix sentence the subordinate clause was more likely to be prosodically integrated as well, i.e. the matrix sentence and the subordinate clause were contained within one prosodic phrase and accordingly covered by one unified phrasal intonation contour. Conversely, if it was syntactically autonomous the subordinate clause also tended to constitute a separate prosodic phrase.

5. Conclusion

A corpus of natural sounding non-scripted standard Danish speech has been created. It is primarily intended for acoustic and perceptual phonetic investigations although it may be also used in empirical morphological, lexical, syntactic and pragmatic studies. The corpus comes with a search engine. Sound files and text files are available to anyone who wishes to use the corpus for research and may be freely downloaded from http://www.danpass.dk, although one does need a password to open the sound files. The password may be obtained from the author.

6. Acknowledgements

This project would not have been possible without extensive help from many people, and not without

external funding either. First and foremost, I am grateful to The Carlsberg Foundation for a grant which permitted me to the hire the student assistants.

A number of individuals have each contributed invaluable help: Preben Dømler and Svend-Erik Lystlund assisted with the recordings. John Tøndering transcribed orthographically all the monologues. He has written a number of immensely useful Praat-scripts for me, for locating mistakes, moving boundaries, etc. He has used the corpus for his own Ph.D. project and liberally shared his results with me. Gert Foget Hansen segmented a part of the monologues. Peter Juel Henrichsen supplied the POS-tagging and is responsible for the search engine. Maja Dyrby and Line Burholt Kristensen proof-read the POS-tags. Nicolai Pharao supplied the 2121 word forms with a phonological representation. Line Burholt Kristensen and Tina Ringkjær added focus and topic tags to the monologues. The major and most tedious work, however, is the responsibility of the assistants who performed the phonetic notation and the prosodic labeling: Cem Avus, Jeppe Beck, Andreas Geisler, Louise Astrid Johansson, Ruben Schachtenhaufen and Thit Wange Stærkær. Finally, without the 27 speakers who gave liberally of their time and enthusiasm, none of this would have been possible.

I am grateful to Michael Fortescue for having taken the time to proof-read the manuscript.

References

Anderson, A.H., Bader, M., Bard, E.G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H.S., Weinert, R. 1991. The HCRC Map Task Corpus. Language and Speech 34, 351-366. See also http://www.hcrc.ed.ac.uk/maptask/.

Boersma, P. 2001. Praat, a system for doing phonetics by computer, Glot International 5:9/10, 341-345.

Boersma, P. and Weenink, D. 2006. Praat: doing phonetics by computer (Version 4.4.30) [Computer program]. Retrieved August 28, 2006, from http://www.praat.org/.

Brown, G., Anderson, A., Shillcock, R., Yule, G. 1984. Teaching Talk. Cambridge University Press, Cambridge.

Dau, T. and Christiansen, T.U. 2007. Byggestene i spontant talt dansk akustisk og perceptuelt. In Kyhn, G. (ed.) Carlsbergfondet. Årsskrift, 44-49.

Dyrby, M. Vinther, Lydiksen, E., Kristensen, L. Burholt, Rørvig, K. 2005. Prosodi i underordnede sætninger i dansk talesprog. Power-point presentation at a meeting in the ProGram circle, November 11th, 2005.

Fletcher, J., Sterling, L., Mushin, I., Wales, R. 2002. Intonational Rises and Dialog Acts in the Australian English Map Task. Language and Speech 45, 229-253.

Grønnum, N. 1985. Intonation and text in Standard Danish, Journal of the Acoustical Society of America 77, 1205-1216.

Grønnum, N. 1986. Sentence intonation in textual context – supplementary data, Journal of the Acoustical Society of America 80, 1040-1047.

Grønnum, N. 1990. Prosodic parameters in a variety of regional Danish standard languages, with a view towards Swedish and German. Phonetica 47, 182-214.

Grønnum, N. 1995. Superposition and subordination in intonation – a nonlinear approach. In Elenius, K., Branderud, P. (Eds.) Proceedings of the XIIIth Internation Congress of Phonetic Sciences, Stockholm 1995, vol. II. KTH and Stockholm University, Stockholm, 124-131.

Grønnum, N. 2005. Fonetik og Fonologi, 3. udg., Akademisk Forlag, København.

Grønnum, N. 2006. DanPASS - A Danish Phonetically Annotated Spontaneous Speech Corpus. In Calzolari, N., Choukri, K., Gangemi, A., Maegaard, B., Mariani, J., Odijk, J., Tapias, D. (Eds.), Proceedings from the 5th International Conference on Language Resources and Evaluation, Genova 24-24 May 2006 (CD-ROM).

Grønnum, N. and Basbøll, H. 2007. Danish Stød – Phonological and Cognitive Issues. In Solé Sabater, M.-J., Beddor, P.S. and Ohala, M. (Eds.) Experimental approaches to Phonology. Oxford University Press, Oxford, 192-206.

Grønnum, N. and Tøndering, J. 2007. Question intonation in non-scripted Danish dialogues. Proceedings of the XVIth International Congress of Phonetic Sciences, Saarbrücken August 5-12 2007, 1229-1232.

Helgason, P. 2006. SMTC - A Swedish Map Task Corpus. Working Papers 52. Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics, 57–60.

Henrichsen, P. Juel, 2002. Sidste Års Aviser – Grammatisk opmærkning af et stort dansk aviskorpus. Lambda 27. Institut for Datalingvistik, Handelshøjskolen i København, København.

Horiuchi, Y., Nakano, Y., Koiso, H., Ishizaki, M., Suzuki, H., Okada, M., Naka, M., Tutiya, S., Ichikawa, A. 1999. The Design and Statistical Characterization of the Japanese Map Task Dialogue Corpus. Journal of Japanese Society for Artificial Intelligence 14, 261-272.

Jensen, C. and Tøndering, J. 2005. Choosing a Scale for Measuring Perceived Prominence. In Trancoso, I. (Ed.) Proceedings of Interspeech 2005, September 4-8, Lisbon, Portugal, 2385-2388.

Kohler, K.J. 2005. Timing and Communicative Functions of Pitch Contours. Phonetica 62, 88-105.

Kohler, K.J. 2006. Paradigms in Experimental Prosodic Analysis – From Measurement to Function. In Sudhoff, S., Lenertová, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N., Schließer, J. (Eds.) Methods in Empirical Prosody Research. (= Language, context, and cognition, 3). de Gruyter, Berlin, New York, 123-152.

Kohler, K.J. 2007. Beyond Laboratory Phonology. In Solé Sabater, M.-J., Beddor, P.S. and Ohala, M., (Eds.) Experimental approaches to Phonology. Oxford University Press, Oxford, 41-53.

Paggio, P. 2006. Annotating Information Structure in a Corpus of Spoken Danish. In Calzolari, N., Choukri, K., Gangemi, A., Maegaard, B., Mariani, J., Odijk, J., Tapias, D. (Eds.), Proceedings from the 5th International Conference on Language Resources and Evaluation, Genova 24-24 May 2006 (CD-ROM).

Silverman K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J. 1992. ToBI: A standard for Labeling English Prosody. Proceedings of the International Conference on Spoken Language Processing, 867-870.

Swerts, M. 1994. Prosodic features of discourse units. Technische Universiteit Eindhoven, Eindhoven.

Swerts, M.. and Collier, R. 1992. On the controlled elicitation of spontaneous speech. Speech Communication 121, 463-468.

Terken, J.M.B. 1984. The distribution of pitch accents in instructions as a function of discourse structure. Language and Speech 27, 269-289.

Tøndering, J. (2008). Sammenhængen mellem prosodi og syntaks i dansk spontantale. Ph.D. dissertation, University of Copenhagen.

Legends to figures

Figure 1

Praat screen. See further details in the text.

Figure 2

Distribution of phrase lengths, in terms of the number of stressed syllables in each phrase, in the monologues in the corpus.

Figure 3

Stylized intonation contour slopes in scripted speech as found in: statements (1); wh-questions (2); questions with word order inversion (3); declarative questions (4).

Figure 1

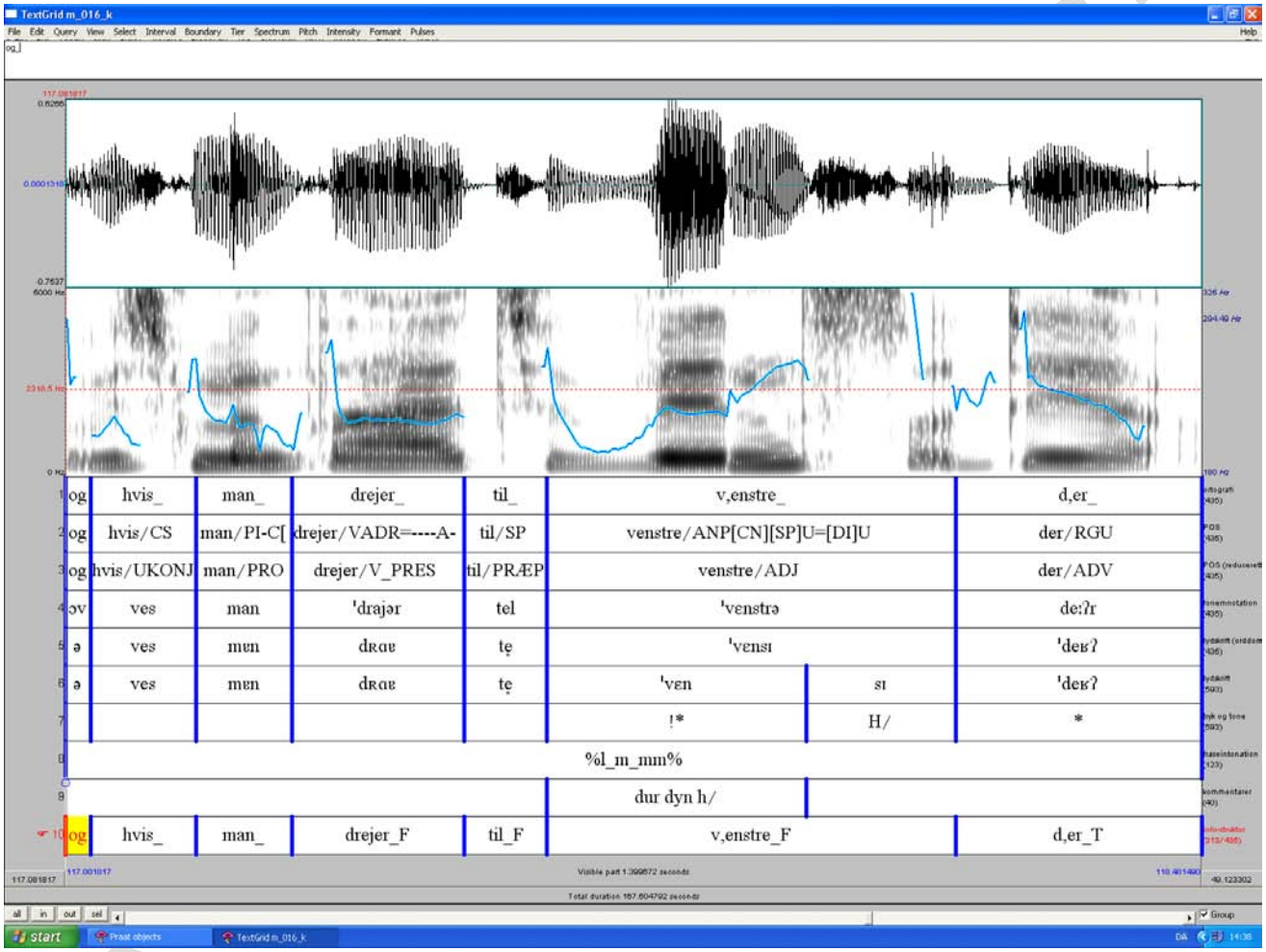


Figure 2

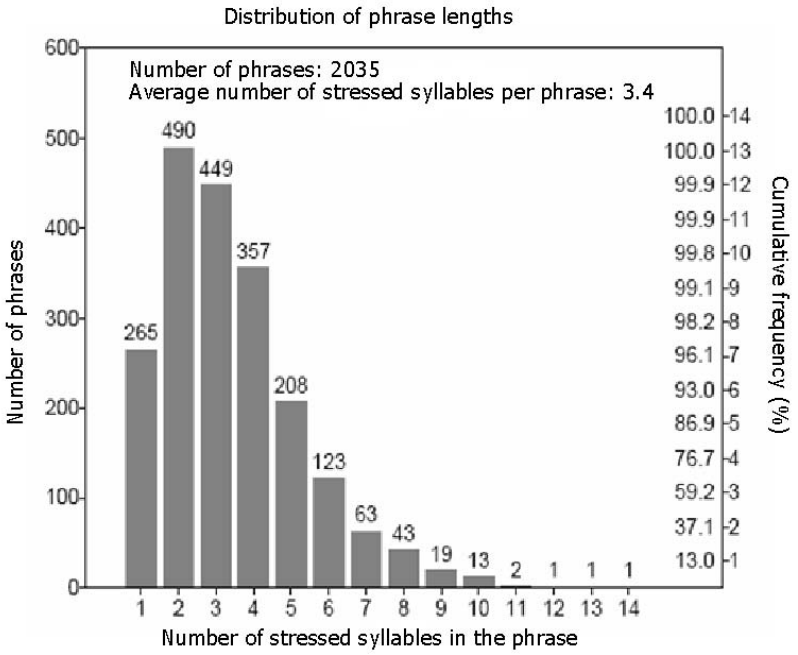
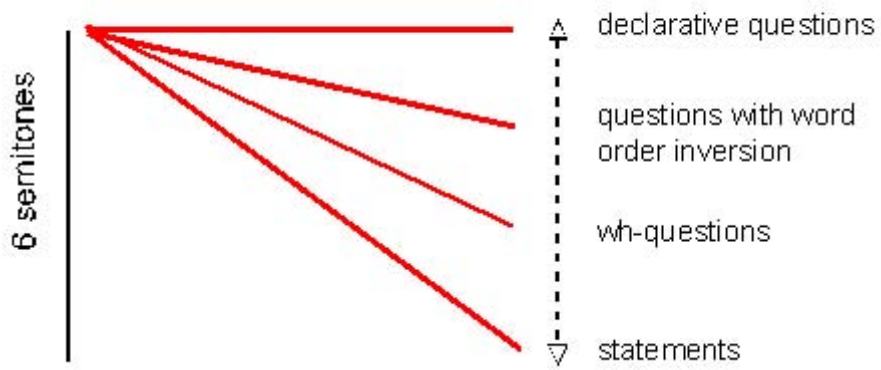
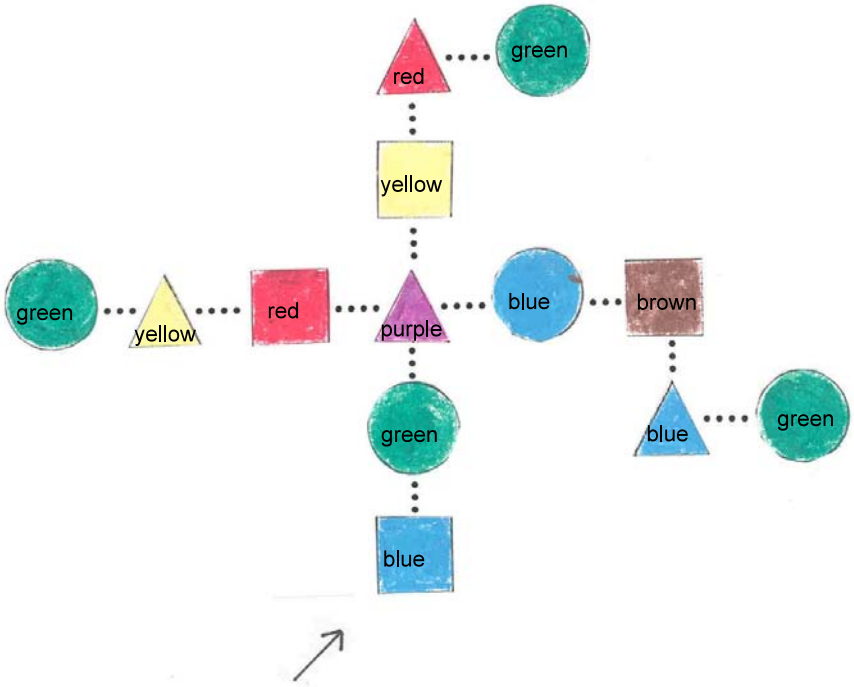


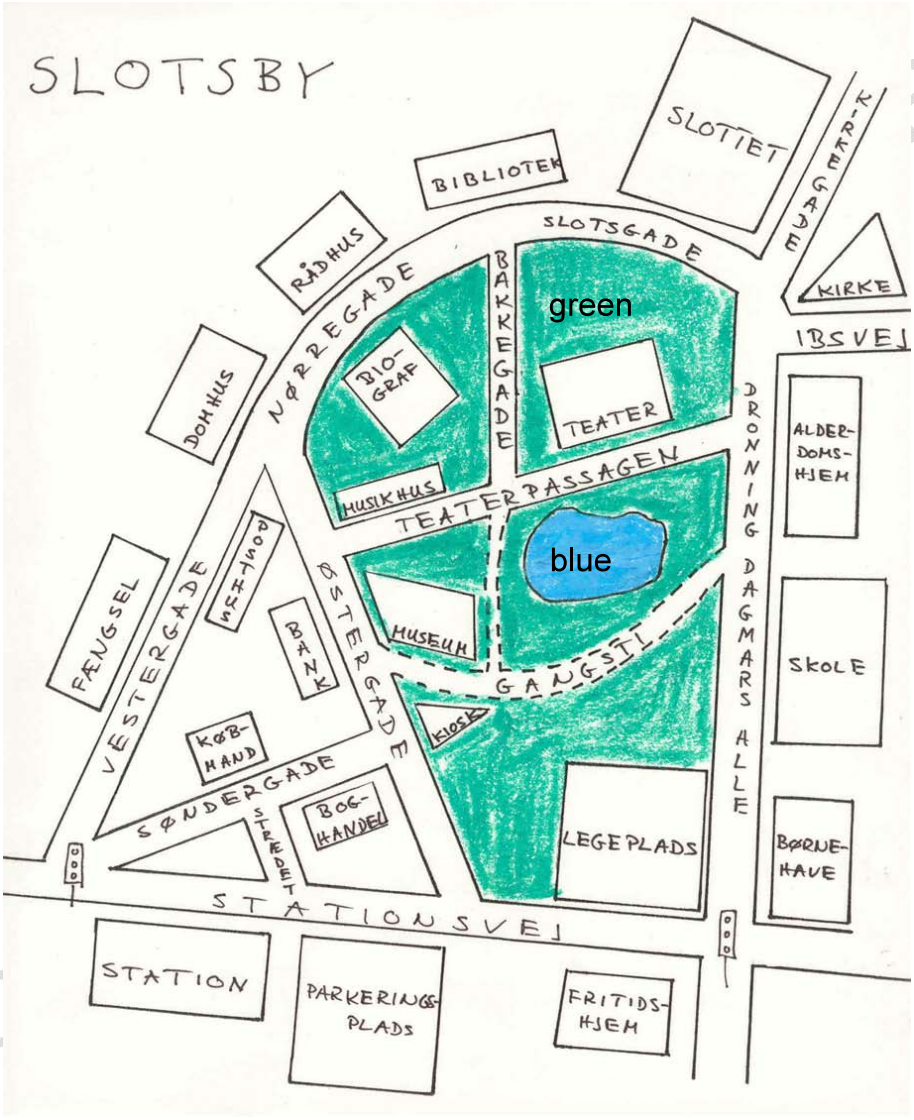
Figure 3



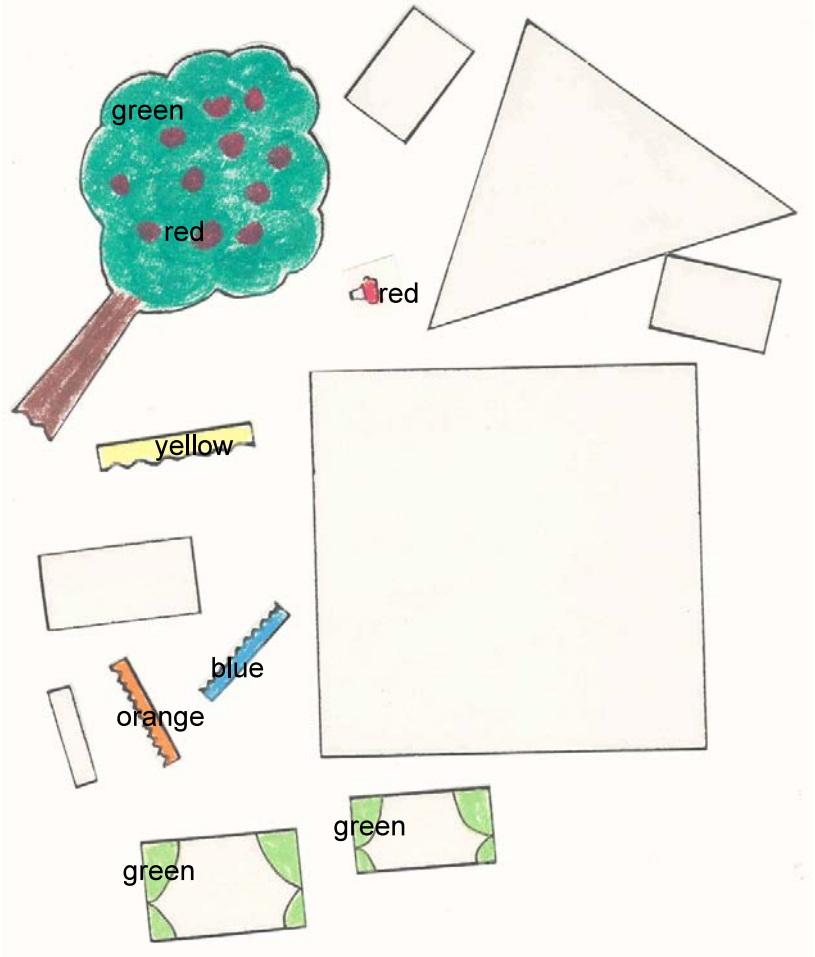
APPENDIX A



APPENDIX B



APPENDIX C



APPENDIX D