LUND UNIVERSITY

# Prosodic Phrasing in Spontaneous Swedish

Hansson, Petra

2003

Link to publication

*Citation for published version (APA):*
Hansson, P. (2003). *Prosodic Phrasing in Spontaneous Swedish*. [Doctoral Thesis (monograph), General Linguistics]. Linguistics and Phonetics.

*Total number of authors:*
1

**General rights**
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

**Take down policy**
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Download date: 05. Sep. 2026

TRAVAUX DE L'INSTITUT DE LINGUISTIQUE DE LUND 43

# Prosodic Phrasing in Spontaneous Swedish

Petra Hansson

LUND UNIVERSITY

Department of Linguistics and Phonetics
Helgonabacken 12
SE-223 62 Lund

© 2003 Petra Hansson

ISSN 0347-2558
ISBN 91-628-5569-7

Photograph (*Skånsk vy*) by Maria Hansson

Printed in Sweden
Studentlitteratur
Lund 2003

# Contents

**Acknowledgments** 8

**Chapter 1: General introduction** 9
1.1 Defining prosodic phrasing 9
1.2 Defining the 'prosodic phrase' 12
1.3 Prosodic phrasing in Swedish 15
1.3.1 Kerstin Hadding-Koch's work on Southern Swedish intonation 15
1.3.2 Eva Gårding's work on prosodic phrasing 16
1.3.3 The research project *Prosodic Phrasing in Swedish* 17
1.4 Prosodic categories in Swedish and their annotation 18
1.5 Spontaneous speech 25
1.6 Aims of the study 27
1.7 Outline 28

**Chapter 2: Phrase boundary distribution** 30
2.1 Introduction 30
2.1.1 Purpose 31
2.1.2 Optimality theory 31
2.1.3 Previous studies 31
2.2 Method 33
2.2.1 Speech material 33
2.2.2 Prosodic transcription 34
2.3 Empirical analysis 34
2.3.1 Speech repairs, Align-XP,R and Wrap-XP 34
2.3.2 Root sentences, Align-XP,R and Wrap-XP 38
2.3.3 Constraints on accentual content and size 39
2.3.4 Maximal focus projection and Align-Focus,R 42
2.4 Summary 44

**Chapter 3: Articulation rate in boundary signaling** 46
3.1 Introduction 46
3.1.1 Phrase-final lengthening in Swedish 47
3.1.2 Interpretations of final lengthening 47
3.1.3 Dialectal variation 48
3.1.4 Articulation rate variation 52
3.1.5 Research question 53
3.2 Method 53
3.2.1 Speech material 53
3.2.2 Segmentation criteria and measurements 55
3.3 Results from the analysis of a subpart of the data 57
3.3.1 Discussion 60
3.4 Results after re-segmentation and addition of more data 61
3.5 Summary and discussion 64

**Chapter 4: Tonal coherence within the prosodic phrase** 67
4.1 Introduction 67
4.1.1 Coherence within the prosodic phrase 68
4.1.2 Time-dependent declination 70
4.1.3 Downstep 72
4.1.4 F0 downtrend in spontaneous speech 76
4.1.5 Research questions and hypotheses 79
4.2 Method 80
4.2.1 Speech material 80
4.2.2 Labeling of prosodic phrase boundaries 80
4.2.3 Measurements 81
4.3 Results 85
4.3.1 Slope and phrase length 85
4.3.2 Slope and F0 starting point 86
4.3.3 F0 starting point and phrase length 89
4.4 Discussion 89
4.4.1 Observations on the qualitative behavior of downstepping accents in Southern Swedish 91
4.4.1.1 Triggering of downstep 92
4.4.1.2 L scaling 93
4.4.1.3 Prominence relations in downstepped sequences of accents 97
4.4.2 Summary 100

**Chapter 5: Tonal coherence among prosodic phrases** 102
5.1 Introduction 102
5.1.1 Phrasal downstep and tonal coupling 103
5.1.2 Research question 106
5.2 Method 106
5.2.1 Speech material 106
5.2.2 Procedure and measurements 107
5.3 Results and discussion 108
5.3.1 The domain of phrasal downstep 108
5.3.2 Signals of coherence vs boundary signals 116
5.3.3 Implications for preplanning 123
5.4 Summary 124

**Chapter 6: Boundary strength** 126
6.1 Introduction 126
6.1.1 Perception of prosodic phrasing in Swedish 126
6.1.2 Perceived boundary strength and the prosodic hierarchy 129
6.1.3 Research questions 138
6.2 Experiment I: Method 139
6.2.1 Stimuli 139
6.2.2 Listeners' task 140
6.2.3 Visual analogue scale 141
6.2.4 Listeners 142
6.2.5 Measurements and normalization 142
6.3 Experiment I: Results and discussion 144
6.3.1 General characteristics of the boundaries in the stimuli 144
6.3.2 The four listener groups 144
6.3.3 Relationship between perceived boundary strength and pause length, F0 reset and final lengthening 146
6.3.4 Relationship between perceived boundary strength and syntactic boundary type 148
6.3.5 Summary 150
6.4 Experiment II: Method 152
6.4.1 Stimuli 152
6.4.2 Listeners and task 153
6.5 Experiment II: Results and discussion 154
6.5.1 Relationship between perceived boundary strength and syntactic boundary type 158

6.5.2 Summary 159
6.6 Discussion 159
6.6.1 Implications for the number of phrasal categories in spontaneous Swedish 159
6.6.2 Phrasal structure in spontaneous Swedish 161

Chapter 7: Summary 163

References 168

Till mormor och morfar
Lisa och Göthe Steinwald

# Acknowledgments

First of all, I would like to thank Inger Enkvist and David House for encouraging me to take on the challenge of writing a thesis. Without your support, I can honestly say that I would never had found the courage needed to do so, and I would have missed out on the rewarding experience it turned out to be.

Next, I wish to express my sincere gratitude to my supervisor, Gösta Bruce, for many stimulating and insightful discussions. His interest in my work and careful manner of presenting constructive criticism created a research environment in which I always felt comfortable pursuing my own ideas.

There are a number of teachers and colleagues that I wish to thank for having inspired me and guided the development of this study. Dawn Behne, Wim van Dommelen, Valéria Molnár and Finn Egil Tønnessen all gave courses that inspired me and influenced my work. I also benefited greatly from discussions with colleagues at the Department of Linguistics and Phonetics in Lund, in particular, from discussions with Eva Gårding, Lars-Åke Henningsson, Anastasia Karlsson, Per Lindblad, Jan-Olof Svantesson and Joost van de Weijer. I am also grateful for important contributions to my work made by Åsa Conway, Maria Mörnsjö, Christina Samuelsson and Paul Touati, and for valuable comments from colleagues at IAAS in Copenhagen and from the special interest group DDISP.

A special thanks goes to my three favorite linguists and phoneticians Paula Kuylenstierna, Mechtild Tronnier and Elisabeth Zetterholm (names in alphabetical order!). More than anything else, I would like to thank you for making these last four years such a good time!

Working together with Merle Horne on the *Swedish Dialogue Systems* project was stimulating and taught me much, as did the collaboration with the other project members. The project was sponsored by a grant from HSFR/NUTEK, which also made this study possible.

I owe a considerable debt of gratitude to the listeners who participated in the perception experiments. My sincere thanks to all of you! I also wish to thank the students whom I had the privilege of teaching in a course on prosody. Your many, many questions and comments made me rethink and reevaluate numerous aspects of prosody, and this study has benefited greatly from that.

For solving countless practical matters, I thank Johan Dahl, Birgitta Lastow, Ingrid Mellqvist and Britt Nordbeck.

Finally, I am deeply indebted to my friends and family for so patiently putting up with me – and numerous phonetic experiments – these last four years. A special thanks to you, Ronnie, for all your support and the unfailing help you provided. For offering me friendship of that valuable kind which can endure long, work-intensive periods of phone silence, I am especially grateful to Johanna and Ingrid. I also wish to express my appreciation to David for his support and patience. And last but by no means least, I wish to acknowledge the love and indispensable support of grandma and granddad, mom and Ulf, Maria and Per, and Tobias.

CHAPTER 1

# General introduction

## 1.1 Defining prosodic phrasing

The stream of speech is interrupted by short pauses or breaks. Speakers group speech into units comprising a handful of words, and in the boundaries between these units or chunks, we hear breaks. In what follows, we will refer to these breaks or boundaries as 'prosodic phrase boundaries'. They are the full stops and commas of spoken language. The division of speech into chunks or 'prosodic phrases' is one of prosody's most important functions, and the topic of the present study.

Despite the great importance generally ascribed to the phrasing function of prosody, numerous researchers have reported difficulties in identifying phrases and phrase boundaries as well as in giving a precise definition of the prosodic phrase (see e.g. Crystal 1969, Gårding and House 1985, Harris, Umeda and Bourne 1981, Liberman 1975, Tench 1995, Umeda and Quinn 1981). Ladd (1986 and 1996) isolates what he believes to be the reason for the often-reported difficulty in defining and identifying phrases. Firstly, he argues that phrase boundaries are not, as often claimed, associated with elusive and hardly audible boundaries, because if they were, "then much of the point of the chunking function would be lost" (Ladd

1996: 235)[1]. The non-elusive character of phrase boundaries is reflected by the high inter-transcriber agreement on the locations of boundaries within transcription systems such as ToBI for English (Pitrelli, Beckman and Hirschberg 1994), GToBI for German (Grice, Reyelt, Benzmüller, Mayer and Batliner 1996), GlaToBI for Glasgow English (Mayo, Aylett and Ladd 1997) and the base prosody system for Swedish (Strangert and Heldner 1995a and b). Even non-expert listeners have been reported to demonstrate good agreement in identifying phrase boundaries (Strangert and Heldner 1995a, Sanderman 1996). The problems Ladd (1996) identifies as related to prosodic or intonational phrases' elusiveness are caused by the internal prosodic structure that is often assumed alongside the presence of audible boundaries (in many cases a so-called 'single most prominent point'). Theoretically incompatible observations such as phrases associated with audible boundaries but without the expected internal prosodic structure and vice versa are a consequence of the potentially conflicting criteria. The same conclusion is drawn in Crystal (1969). What is regarded an unexpected internal structure is theory-dependent; it may be a structure that lacks a 'designated terminal element' (Liberman and Prince 1977), a 'nuclear accent' (Cruttenden 1986) or a 'phrase accent' and a boundary tone (Pierrehumbert 1980).

One possible way to characterize prosodic phrasing is by describing how it is used in speech. Prosodic phrasing as a cue to syntactic structure has been investigated by a large number of researchers, although predominantly in laboratory speech. Cutler, Dahan and van Donselaar (1997) review a large number of studies on the role of prosody in the computation of syntactic structure undertaken in the 60's, 70's, 80's and 90's[2]. They conclude that the presence of a prosodic boundary indeed can have an effect on syntactic analysis.

That prosody has the potential to aid the understanding of syntactic structure is supported by studies showing that listeners can accurately locate major syntactic boundaries from prosody alone (Collier and 't Hart 1975) and by studies testing the comprehension of differently phrased utterances (Sanderman and Collier 1997). Investigations of sentences that are globally ambiguous, i.e. sentences with ambiguities that are not resolved by the occurrence of further linguistic (non-

[1] Although some researchers have proposed that phonological domains such as the prosodic phrase need not be defined with reference to its boundaries (see Ladd 1986), we will not choose to do so here as we feel that such a proposal is not compatible with the view that prosodic phrasing plays a role in chunking.

[2] More recent work on the relationship between prosody and syntax can be found in e.g. Kang and Speer (2002) and Jun (2002).

prosodic) information within the sentence, give additional support to the hypothesis that listeners can use prosody to understand what structural interpretation of a sentence the speaker intended (Lehiste 1973, Bruce, Granström, Gustafson and House 1993). Studies of sentences with local ambiguities, i.e. that are disambiguated as the sentence unfolds (by the occurrence of further linguistic information), also suggest that listeners use prosodic information to resolve ambiguities in the structural interpretation of the sentence. One example of such a study was undertaken by Grosjean (1983) who showed that listeners are able to predict utterance length (defined as the number of prepositional phrases following a verb phrase: *Earlier my sister took a dip / in the pool / at the club / on the hill*) using prosodic information alone. However, Cutler *et al.* (1997) also note that the effects of prosody on syntactic analysis are far from robust and determinative, and that little support has been found to suggest that prosody is used early in processing. Although prosody clearly can matter in syntactic parsing, its exact role is not yet clear. Prosody can be seen as a linguistic structure providing information to the parser on its own, or as a provider of potentially ambiguous information that the parser can use. In a series of perception experiments with Swedish stimuli, House (1985) tests the hypothesis that the use of prosodic information in perception decreases as more syntactic information is available from lexical redundancy rules or morphological restrictions. Once again, results indicate that prosody can be used in speech perception to parse a sentence. However, House (1985) also concludes that prosodic cues are not always needed for correct parsing and, conversely, that prosody is not always used. The results are interpreted as evidence for a model of syntactic parsing that operates by simultaneously integrating prosody, syntactic complexity strategies and morphological restrictions in the lexicon.

Cutler *et al.* (1997) relate some of the problems associated with the use of prosody in the computation of syntactic structure to the non-isomorphic relationship between syntactic and prosodic structure (see Shattuck-Hufnagel and Turk 1996 for a discussion). The fact that syntactic and prosodic structure is not isomorphic motivated Chomsky and Halle (1968) to design readjustment rules which alter the surface structure to a division into 'phonological phrases'. The idea that the hierarchical structure in syntax coexists with a separate phonological or 'prosodic hierarchy' with constituents that are not necessarily identical to those in the syntactic hierarchy, has subsequently been elaborated by Liberman and Prince (1977) (who nevertheless considered the branching of the trees isomorphic above the word level), Nespor and Vogel (1986), Selkirk (1984) and Beckman and Pierrehumbert (1986).

The hypothesis tested by House (1985) gives an interesting perspective on the so-called elusiveness of prosodic phrase boundaries. If indeed the use of prosodic information in perception decreases as more syntactic information is available, then it may be that prosodic information also decreases in production with increasing syntactic information. The elusiveness of some phrase boundaries is then directly related to their limited potential for being beneficial to the listener. Some boundaries may be "elusive" without the chunking function thereby being lost. In addition, it should be mentioned that the elusiveness of certain phrase boundaries, particular in spontaneous speech, may also be a consequence of the speaker's unawareness of an ambiguity in the message being conveyed (Hirschberg 1999, Lehiste 1973). Finally, chunking is beneficial not only to the listener but also to the speaker who may need the time that pausing and phrase-final lengthening provide for planning the upcoming speech and (in the case of pausing) for breathing. Consequently, many clearly audible phrase boundaries will also be found in positions where their existence makes no necessary contribution to the listener's comprehension of syntactic structure.

Returning to our discussion on the usages of prosodic phrasing in speech, prosodic phrasing is also used to indicate which words within a sentence belong together semantically or pragmatically. In many cases, semantically coherent units of speech are also marked syntactically, and therefore it is often difficult to tease apart the roles that prosody plays in signaling syntactic and semantic information. Differences in how the same syntactic structure is phrased can be observed if semantic weight is taken into consideration. Semantically richer syntactic phrases tend to be longer and, because of their length, form separate prosodic phrases (Bing 1985, Bruce 1998).

## 1.2 Defining the 'prosodic phrase'

In the introduction to a paper on phrase intonation in Swedish, Gårding and House (1985: 205) note that they "as little as anyone else can give a precise definition of the concept [prosodic] phrase" (author's translation). An approximate definition is nevertheless offered to the reader, namely that "a prosodic phrase is a part of the utterance that organizes accents or tones in a common, unbroken intonation movement", i.e. a sequence of adjacent 'prosodic words' with internal tonal cohesion. Of reasons that will be made clear below, we will use this definition of the prosodic phrase in the present study as well, rather than a definition that at the surface may appear more precise, e.g. a definition that rests on the identification

of a phrase accent or boundary tone. Nevertheless, we will try to further define the prosodic phrase by comparing it with other similar prosodic constituents that have been proposed for other languages.

In an overview of prosodic constituents in the literature, Wightman, Shattuck-Hufnagel, Ostendorf and Price (1992) note that the 'intonational phrase' of most intonation models can be defined loosely as a group of words in an utterance that is delimited in some way as a larger unit of phrasing. Ladd (1986) identifies two further properties of the intonational phrase that generally are assumed (in addition to being the largest chunk into which utterances are divided). They are, firstly, a tie to elements of syntactic structure (which was discussed in section 1.1), and secondly, a single most prominent point (e.g. a 'tonic', 'nucleus' or 'phrase accent'). We will discuss the most prominent point of the phrase in further detail below and in section 6.1.2.

The most influencial definition of the intonational phrase is perhaps the definition proposed by Pierrehumbert (1980). The intonational phrase as defined by Pierrehumbert has distinctive tonal characteristics, namely a boundary tone (H% or L%) occurring at the phrase boundary and a phrase accent ($H^-$ or $L^-$) which is placed after the nuclear accent. It thereby resembles Halliday's (1967) 'tone group' in the sense that it is the prosodic group's tonal properties that are foregrounded. The Pierrehumbert (1980) and subsequently also the ToBI definition (Beckman and Ayers 1993, Silverman *et al.* 1992) of the intonational phrase can be argued to be more straightforward than e.g. the definition given by Gårding and House (1985) in the sense that it identifies a phrase edge tone. However, it is not always the case that an intonational phrase boundary is associated with a clearly observable change in the tonal domain. This fact is reflected in the British English system IViE (Grabe, Post and Nolan 2001) where the transcriber has the option to mark the presence of a phrase boundary without associating it with a boundary tone. When the pitch level reached at the end of the last accent in the intonation phrase continues at the same level, no tone is specified.

The prosodic phrase in Scandinavian languages other than Swedish has been described by, among others, Fretheim (1981, 1991 and 2001) and Grønnum Thorsen (1988). Fretheim defines the 'intonational phrase' in (East) Norwegian mainly with regards to its internal accentual structure. It is defined as a constituent comprising one or more 'accent units' or feet (Fretheim 1981 and 1991). The accent unit consists of an accented word (which is pronounced with one of the two opposing word accents) and the following unaccented words (if any). It is either

'attenuated'/'nonfocal' (Fretheim 1981 and 1991) or 'unattenuated'/'focal'/'phrase-accented' (Fretheim 1981, 1991 and 2001). The accent unit is thus similar to the prosodic word in the Swedish intonation model[3]. Although Fretheim allows so-called 'backgrounding' intonational phrases that comprise only attenuated accent units, an unattenuated accent unit always marks the end of an intonational phrase (Fretheim 1981), and thus has a similar status in the Norwegian phrase as e.g. the phrase accent has in the English phrase in Pierrehumbert's (1980) work. Fretheim's definition thus employs the idea of a single most prominent point in the phrase. It is thereby different from the definition of the prosodic phrase in Danish. Grønnum Thorsen (1988) identifies the prosodic phrase as a component that adds a phrasal contour to the sentence intonation contour. It consists of one or several 'prosodic stress groups' (which may be modified by 'stød'). The prosodic stress group is defined in much the same way as the accent unit in Norwegian and the prosodic word in Swedish, i.e. as a stressed syllable and all succeeding unstressed syllables (if any). The Danish phrase is not characterized by a phrase-final accent that is necessarily different from the non-final accents (i.e. by an obligatory 'sentence accent' or 'nucleus'). No stressed syllable is more prominent than the others in a pragmatically neutral utterance (Thorsen 1983). In this regard, Norwegian and Danish are similar to Stockholm and Southern Swedish, respectively. Whereas the prosodic phrase in Stockholm Swedish has a default, phrase-final focal accent (Bruce 1977), Southern Swedish does not.

Beckman and Pierrehumbert (1986) have claimed another level of phrasing between the prosodic word and the intonational phrase, namely the 'intermediate phrase'. This intermediate phrase is similar to the 'phonological phrase' as defined by Nespor and Vogel (1986) and the 'major phrase' as defined by Selkirk (1984). Its hallmark is the presence of a phrase accent. As will be discussed in the next section and subsequently also in chapter six, the intermediate phrase is generally not regarded as a relevant phrasal category in Swedish.

As regards a higher-level phonological constituent, a level of phrasing above the intonational phrase, categories such as the 'phonological utterance' (Nespor and Vogel 1986) and the 'prosodic utterance' (Bruce 1994) have been suggested in the literature. The phonological utterance has been motivated by the existence of

[3] The 'prosodic word' in Swedish can be defined as a sequence comprising a primary stressed syllable (pronounced with either accent I or II) and the following unstressed syllables (if any). In compounds, it furthermore contains a secondary stressed syllable and the unstressed syllables following it (if any).

phonological rules like flapping in American English. It makes use of syntactic information in its definition (Nespor and Vogel 1986).

Liberman and Pierrehumbert (1984) have found prosodic phenomena with a possibly larger domain than the intonational phrase. The endings of declarative sentences were found to be subject to final lowering. Based on this observation and similar observations from Japanese on declination, a higher-level unit, the 'utterance', was posited in Beckman and Pierrehumbert 1985. Nevertheless, in Beckman and Pierrehumbert 1986, it is declared that more detailed investigations undermine such a claim. The phonetic effects in question can be related to discourse structure. Final lowering is controlled by discourse structure in a manner that makes it implausible to claim that it defines a higher-level phonological constituent. The higher-level constituents 'phonological paragraph' (Lehiste 1975) or 'speech paragraph' (Bruce 1994) are examples of other constituents posited and subsequently omitted from the prosodic hierarchy (Selkirk 1984, Nespor and Vogel 1986, Bruce, Granström, Gustafson, House and Touati 1994). In Grønnum Thorsen (1988), the so-called 'overall textual contour' and 'sentence intonation contour' are referred to as non-categorical components.

## 1.3 Prosodic phrasing in Swedish

### 1.3.1 Kerstin Hadding-Koch's work on Southern Swedish intonation

In her doctoral dissertation *Acoustico-Phonetic Studies in the Intonation of Southern Swedish*, Kerstin Hadding-Koch (1961) investigated the functions of intonation in connected speech in Southern Swedish. Whereas many studies on intonation undertaken since then have focused on the so-called standard variety of Swedish (e.g. Gårding 1967a and Bruce 1977), Hadding-Koch regarded Southern Swedish to be a convenient object to try out various approaches to intonation analysis. It should be noted that Hadding-Koch took 'Southern Swedish' to mean *skånska* 'Scanian', the dialect spoken in the southernmost region of Sweden. In the prosodic typology later developed by Gårding and Lindblad (1973) and Bruce and Gårding (1978), 'South Swedish' (or dialect 1a) includes more dialects than those which are spoken in *Skåne* 'Scania'. In the present study, we chose to follow Hadding-Koch in the sense that we will use Southern Swedish (*skånska*) as a suitable object to investigate in our study of a phenomenon in spontaneous speech mainly known to

us from studies of read speech, namely prosodic phrasing. We are thus not intending to describe the Southern Swedish dialect *per se*.

Hadding-Koch's (1961: 189) studies of intonation as an "instrument for expressing syntactical relations between utterances and parts of utterances" are the ones that are most relevant to us. She concluded that intonation is used to express syntactical relations in connected speech by both its function as an important correlate to prominence and as a means to express internal and terminal 'junctures'. As regards the signaling of junctures, she nevertheless noted that other, non-tonal features, such as duration and intensity, are also involved. In the present study, as in most Swedish prosody research, we will acknowledge this fact about junctures in Swedish, and refer to them as 'prosodic' phrase boundaries rather than 'intonational' phrase boundaries.

### 1.3.2 Eva Gårding's work on prosodic phrasing

Eva Gårding distinguishes between what she terms 'internal junctures' and 'terminal junctures'. Her doctoral dissertation *Internal Juncture in Swedish* (1967a) deals with perceptual, acoustic and articulatory aspects of internal juncture in Swedish. The internal juncture is defined as a marked syllable boundary in a phrase. Minimal pairs like *lätta tankar – lättat ankar* ('simple thoughts' – 'weighed anchor') demonstrate the internal juncture's ability to change the meaning of a phrase. It occurs at word and morpheme boundaries between consecutive stressed vowels, and can easily be recognized by listeners because of the glottal closure (when a vowel follows the juncture) or aspiration (when a consonant follows) that arise as the speech organs slow down and move toward a neutral position. As regards the prosodic boundaries of interest in the present study, Gårding (1967a) follows Hadding-Koch (1961) and terms them terminal junctures.

Gårding and colleagues' work on terminal junctures or prosodic phrasing includes studies of both production (Gårding 1974, Gårding and House 1985, Gårding and House 1986) and perception (Gårding and Eriksson 1989, Gårding and House 1985, Gårding and House 1986). We will discuss some of these studies in further detail below, e.g. Gårding's investigations of stress patterns within the phrase in Stockholm and Southern Swedish (see section 3.1.3). In chapter four, we will furthermore review some of the features of the Lund intonation model for intonation as advocated by Eva Gårding. The Lund model for intonation has been revised (Bruce 1982a, 1982b and 1984) and therefore exists in two versions: Gårding's hereafter termed 'original' version of the Lund model and Bruce's

'revised' version. A comparison of the two intonation models can be found in Gårding (1987).

### 1.3.3 The research project *Prosodic Phrasing in Swedish*

The research project *Prosodic Phrasing in Swedish*, led by Gösta Bruce and Björn Granström, was a part of the HSFR/NUTEK financed Swedish Language Technology Program 1990-93, and a joint effort between the Department of Linguistics and Phonetics at Lund University and the Department of Speech Communication and Music Acoustics at KTH in Stockholm (Bruce and Granström 1993, Bruce, Granström, Gustafson and House 1991, Bruce *et al.* 1993). The work in Lund was aimed at developing the intonation model for Swedish, and the work in Stockholm towards the development of the prosodic component in a text-to-speech system. The research questions addressed concerned both the phonology and the phonetics of prosodic phrasing. The main phonological issue was to gain knowledge about what types of prosodic phrases are relevant domains in Swedish, more specifically, whether it is relevant to speak about an intermediate phrase as a relevant domain between the prosodic word and the prosodic phrase in Swedish. The main phonetic issue concerned what speech variables and combination of variables are used to signal coherence within and boundaries between prosodic phrases, both locally and globally. Three methods were used in the investigation: 1) analyses of read production data, 2) text-to-speech synthesis and 3) speech recognition (prosodic parser).

The main conclusions relevant to the development of the intonation model of Swedish were concerned with the phonetics of prosodic phrasing in Swedish. The production data studies revealed that several different phrasing strategies (different combinations of F0 and duration cues contributing to coherence and boundary signaling) were exploited to disambiguate sentences. A series of perception tests gave further insight into the use of F0 and duration cues in the perception of phrasing. Results revealed that most listeners rely on a combination of F0 and duration cues, although primarily "duration-minded" and "F0-minded" subjects were also reported to exist (Bruce *et al.* 1993). No evidence to support the 'intermediate phrase', as defined in Pierrehumbert (1980), as a relevant phrasal category in Swedish was reported, and consequently the intermediate phrase was not included in the Swedish intonation model.

No investigations of prosodic phrasing in spontaneous speech were undertaken, as the work done within the project represented a return to studies of laboratory speech and highly controlled conditions. Furthermore, the only variant of Swedish examined was the so-called standard variety (dialect 2a in Bruce and Gårding's (1978) prosodic typology).

## 1.4 Prosodic categories in Swedish and their annotation

Two transcription systems have been specifically developed for the annotation of prosody in Swedish: the IPA-based 'base prosody system' and the ToBI-like 'tonal transcription system'. A review of the two systems will reveal what the phonological domains and categories are in the Swedish intonation model.

The base prosody system for Swedish, an IPA-based transcription system, was developed within the HSFR/NUTEK financed Swedish Language Technology Programme 1990-96. It is the result of a national discussion among phoneticians specializing in prosody and a proposal for a common system for transcribing Swedish prosody. Transcriptions made with the base prosody system rely on an auditory analysis, and are meant to be phonological rather than phonetic. The base prosody system contains symbolization of the categories prominence and grouping (or boundary phenomena) on a phonological level (Bruce 1994). Below, the symbols of the base prosody system are given.

Prominence categories:

| | | |
|---|---|---|
| ˈˈcv | Focused or focally accented, accent I | Extra strong prominence |
| ˈˈcv̀ | Focused or focally accented, accent II | Extra strong prominence |
| ˈcv | Primary stressed or accented, accent I | Strong prominence |
| ˈcv̀ | Primary stressed or accented, accent II | Strong prominence |
| ˌcv | Secondary stressed | Weak prominence |
| | Unstressed | No marking |

Boundary categories:

| | | |
|---|---|---|
| cv ‖‖ cv | Extra strongly marked boundary | Corresponding to e.g. speech paragraph |
| cv ‖ cv | Strongly marked boundary | Corresponding to e.g. prosodic utterance |
| cv \| cv | Weakly marked boundary | Corresponding to e.g. prosodic phrase |
| cv cv | No boundary | No marking |

*'cv' refers to any syllable, with 'c' and 'v' representing the consonant and vowel, respectively.*

(Based on Bruce 1994: 15)

Representations of three prominence categories (in addition to the unstressed condition) are included in the base prosody system: (secondary) stress, (primary stress or non-focal) accent and (focus, sentence accent[4] or) focal accent. It is important to note that the three prominence levels assumed do not only reflect perceptually distinguishable degrees of prominence but also three communicatively relevant categories. The motivation for assuming three prominence categories or three communicatively relevant degrees of prominence can be demonstrated with minimal pairs.

The communicative relevance of the categorical dichotomy between unstressed and (secondary) stressed can be demonstrated with a minimal pair like ˈ*dànskorna* 'the Danish women' – ˈ*dàns*ˌ*skorna* 'the dancing shoes' (example from Bruce 1977: 13). The presence or absence of a (secondary) stress, in this particular case on the syllable 'skor', is distinctive. The secondary stress' placement is also distinctive as illustrated by the minimal pair ˈ*nàcka*ˌ*schacket* 'the Nacka chess' – ˈ*nackaja*ˌ*kett* 'Nacka morning coat' (examples from Bruce 1977: 14).

[4] In early work by Bruce (1977), the 'focal accent' (Bruce 1987) is termed 'sentence accent'. The accent in question has also been referred to as a 'phrase accent' (Pierrehumbert 1980).

The motivation for separating stress from accent can be shown by contrasting a two-word prosodic phrase such as *mellan målen* 'between the meals' and the segmentally identical one-word prosodic phrase *mellanmålen* 'the snacks' (example from Bruce 1998: 140), see Figure 1.1. Whereas *mellan* carries an accent in both phrases, *målen* is only accented in the two-word phrase. In other words, the difference is one of accent. In the one-word phrase, the first syllable of *målen* is (secondarily) stressed, but not associated with an accent and therefore perceived as part of the same word as *mellan* (Bruce 1977). Perceptual evidence of the distinction between stress and accent can be found in Zetterlund, Nordstrand and Engstrand (1978) and Gårding and Eriksson (1989).

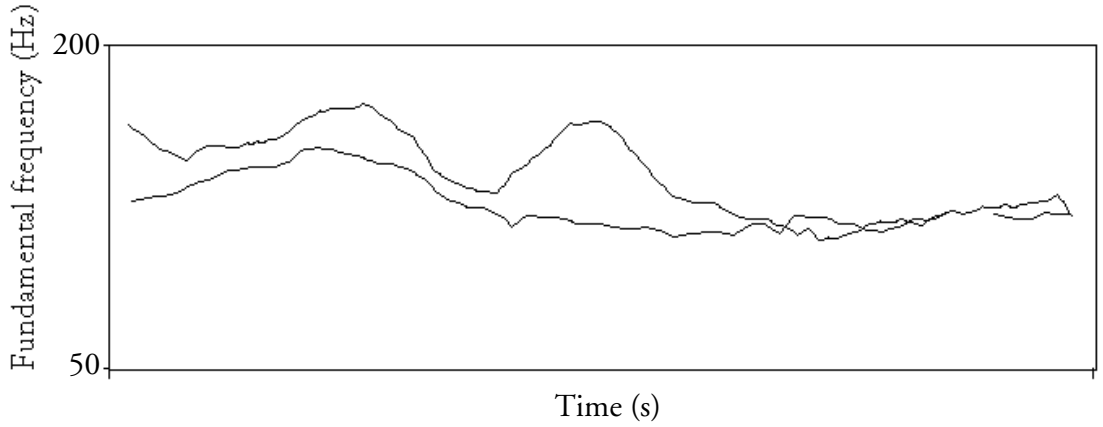


**Figure 1.1** *F0 contours of* mellan målen *'between the meals' (top line) and* mellanmålen *'the snacks' (bottom line) (male Southern Swedish speaker).*

Swedish is a language with a lexically and morphologically conditioned distinction of accent type. As described above, the primary stressed syllable is associated with an accent. The accent is either acute (accent I) or grave (accent II) and will hereafter be termed 'word accent'[5]. In some contexts in this study, it will be contrasted with the 'focal accent'[6] (the third degree of prominence), and therefore also be referred to as a 'non-focal accent'. It is important to note that there is no difference in degree of prominence between accent I and II. Rather, they are phonological properties of individual word forms. Phonetically, the difference between them is one of F0 peak timing. In all dialects of Swedish[7], the F0 peak of accent I has an earlier alignment with the stressed syllable than accent II (Bruce and Gårding 1978, see also Malmberg 1963). In Figure 1.2, an example is given of the peak timing in an accent I- and II-word in one Swedish dialect, namely Stockholm Swedish.

[5] In the literature, it is also termed 'pitch accent'.

[6] The word accent distinction is maintained also in focal position.

[7] Except in Finland Swedish, where the word accent distinction is not maintained (Bruce and Gårding 1978).

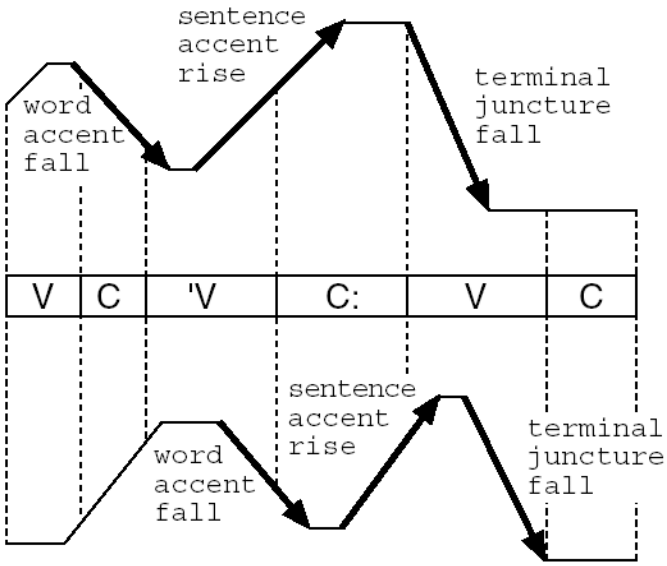


**Figure 1.2** *Schematized F0-contours of one accent I- and one accent II-word in Stockholm Swedish (From Bruce 1977: 50, © Gösta Bruce 1977. Reprinted with permission).*

In the same way as the secondary stress' placement is distinctive, so is the placement of the primary stress and thereby the word accent. In other words, segmentally identical words can be categorically distinct not only when they differ in word accent type (e.g. ˈ*anden* 'the duck' – ˈ*ànden* 'the spirit') but also in word accent placement (ˈ*formel* 'formula' – *for*ˈ*mell* 'formal').

Moving on to the next level of prominence in the model, the motivation for distinguishing between accent and focal accent can be illustrated by question-answer pairs like those in (1a)[8]. Prosodically illformed answers are marked with a star (see Bruce 1977: 21-24 for a discussion). In the English translation, focally accented words are written with capital letters.

(1a)
–Vad är det för gula blommor? 'What kind of yellow flowers are those?'
–ˈGùla tulˈˈpaner. / *ˈˈGùla tulˈpaner. 'Yellow TULIPS. / *YELLOW tulips.'
– Vad är det för tulpaner? 'What kind of tulips are those?'
– *ˈGùla tulˈˈpaner. / ˈˈGùla tulˈpaner. '*Yellow TULIPS. / YELLOW tulips.'

(Bruce 1998: 83)

[8] That the relevant difference is between accent and focus (and not between stress and accent as in the examples in Figure 1.1) is corroborated by the fact that the word accent distinction is maintained in non-focal position.

Perceptual evidence of the distinction between accent and focus is reported on in Horne and Filipsson (1998). Perceptual testing of text-to-speech systems with and without a referent tracker (a component in the linguistic preprocessor that recognizes contextual coreference or cospecification relations between content words based on givenness, morphological identity and lexical semantic identity-of-sense relations) indicated that listeners prefer a system that includes a referent tracker to one that does not. When asked if they preferred 1) *Den ''ròda 'bilen är min favo''rit* 'The RED car is my FAVORITE' or 2) *Den ''ròda 'bilen är min favo'rit* 'The RED car is my favorite' as an answer to the question *Vilken ''bil 'tycker du ''bäst om?* 'What car do you like best?', 79% of 94 listeners preferred answer (2). In other words, they preferred the answer where only the 'new' information was focally accented. The relationship between focal accentuation and information structure will be discussed in further detail below.

Further differences in degree of prominence can be perceived by listeners, e.g. different degrees of emphasis, but they are regarded as variation within the phonological categories discussed above. In the case of emphasis, perceivable differences are regarded as variation within the category focus or focal accent. In the implementation of the intonation model in Bruce and Granström (1993), eight different degrees of phonetic emphasis are assumed within the focus category.

In the base prosody system, a distinction is also made between three boundary types and thereby three types of phrasal categories: prosodic phrases (which are delimited by weak boundaries indicated with '|'), prosodic utterances (which are delimited by strong boundaries indicated with '||') and speech paragraphs (which are delimited by extra strongly marked boundaries indicated with '|||'). The relationship between the number of boundary strengths acknowledged and the number of phrasal categories assumed will be discussed in further detail in chapter six. An evaluation of the base prosody system has been undertaken by Strangert and Heldner (1995a and b).

The second, so-called tonal transcription system for Swedish is a system "not unlike ToBI" (Bruce *et al.* 1994: 36) which was developed within the research project *Prosodic Segmentation and Structuring of Dialogue*, a joint effort between the Department of Linguistics and Phonetics in Lund and Department of Speech Communication and Music Acoustics at KTH, the Royal Institute of Technology, in Stockholm, and a part of the HSFR/NUTEK financed Swedish Language Technology Programme 1993-96.

The tonal transcription system is dialect-specific and was developed for the so-called standard variety of Swedish. It contains the same tonal prominence categories as the base prosody system, i.e. accent and focal accent. Stress (defined as the first degree of prominence in the model) has no tonal correlate in Swedish (Bruce 1977).

Instead of an abstract symbolization such as that used in the base prosody system, the labels used in the tonal transcription system reflect the F0 pattern typically associated with the categories in question. Transcriptions made rely to some extent on an acoustic-phonetic analysis, although the auditory analysis is by no means unimportant. Despite the labels' reference to the categories' phonetic form, the tonal transcription is also meant to be phonological. Below, the symbols of the tonal transcription system are given.

Tonal categories:

| | | |
|---|---|---|
| (H)L*H[9] | Focal accent, accent I | Extra strong prominence |
| H*LH | Focal accent, accent II | Extra strong prominence |
| HL* | Primary stress or accent, accent I | Strong prominence |
| H*L | Primary stress or accent, accent II | Strong prominence |
| %L | Low initial boundary tone | *Used in combination with the boundary category labels below* |
| %H | High initial boundary tone | |
| L% | Low final boundary tone | |
| H% | High final boundary tone | |

Boundary categories:

| | | |
|---|---|---|
| cv \|\| cv | Strongly marked boundary | Corresponding to e.g. prosodic utterance |
| cv \| cv | Weakly marked boundary | Corresponding to e.g. prosodic phrase |
| cv cv | No boundary | No marking |

*H and L refer to high and low tones, respectively. Starred tones are critically associated with the stressed syllable.*

(Based on Bruce 1998: 168)

The term 'focal accent' highlights the accent's function in speech, namely as a marker of the focus of an utterance (see example (1a) above). In what follows, we will briefly discuss the relationship between phonetic prominence (focal accentuation) and focus in information structure. Although focal accentuation and

[9] The phrase accent is sometimes distinguished from the word accent tones by adding a '–'.

phrasing are easily separated in theory, they are interwoven in the practical situation.

A referent that is introduced into the discourse for the first time is said to be 'new' whereas a 'given' referent is one which has been mentioned in the previous discourse or is inferable from known information (Dahl 1976, Jespersen 1924, Strawson 1964). In some related previous studies (see e.g. Horne and Filipsson 1998), the 'new'/'given' status of individual words has been assumed to play an important role for the distribution of focal accents within the utterance. It is a somewhat simplified view (see e.g. Bolinger 1972, Halliday 1967, Terken and Hirschberg 1994), as it is not individual 'new' words *per se* that are assigned focal accents but the utterances' foci (Ladd 1996, Lambrecht 1994). The 'focus' of an utterance is the part that makes the utterance a piece of information. As pointed out by Halliday (1967: 204), the focal information is 'new', "not in the sense that it cannot have been previously mentioned […] but in the sense that the speaker presents it as not being recoverable from the preceding discourse". Thus, as noted by Molnár (1998: 129), adding new information is "not only possible through the introduction of new referents but [also] by the "mere" expression of different types of "new" (unpredictable or not yet settled) relations".

It is generally accepted that focal accentuation reflects the intended focus of an utterance. However, there is disagreement about how focus is conveyed by (focal) accent. According to the 'Focus-to-Accent' (FTA) approach (Gussenhoven 1983), the focus of an utterance is marked by a focal accent. In the case of 'narrow focus' (focus on an individual word), accent goes on the focused word, but in the case of 'broad focus', i.e. focus on whole constituents or whole sentences, language-specific or perhaps even dialect-specific rules are applied that decide which word takes the focal accent (Gårding 1974, Ladd 1996). Since several 'new' words may be contained within the focus constituent, the FTA approach does not predict that all 'new' words be associated with focal accents. Conversely, the focus constituent does not have to contain any 'new' *lexical* information at all and therefore 'given' referents sometimes take a focal accent. Despite of the term's reference to its function as a marker of the utterance's focus, it should be noted that the HLH label is used to annotate the third level of prominence in Swedish, a fall-rise F0 pattern in the case of Stockholm Swedish, regardless of whether the accent in question was used to mark the focus or some other part the utterance, i.e. even for accents which in the literature on information structure may be referred to as 'topic accents'. A discussion on topic accents can be found in Lambrecht (1994).

Finally, in the tonal transcription system, a distinction is made between two types of phrasal categories: prosodic phrases (which are delimited by weak boundaries indicated with '|') and prosodic utterances (which are delimited by strong boundaries indicated with '||'). The third boundary strength in the base prosody system, extra strongly marked boundaries, is not annotated. Unlike the base prosody system, the tonal transcription system has not been evaluated. However, as we feel that the tonal transcription system and the base prosody system complement each other, we will use both in the present study.

No complete guidelines with training materials for either the base prosody nor the tonal transcription system are readily available, as they are for e.g. ToBI and American English (http://www.ling.ohio-state.edu/~tobi/, accessed 2003-01-06), IViE and Brittish English (http://www.phon.ox.ac.uk/~esther/ivyweb/guide.html, accessed 2003-01-06), GToBI and German (http://www.coli.uni-sb.de/ phonetik/, accessed 2003-01-06) and for ToDI and Dutch (http://lands.let.kun.nl/todi/todi/ home.htm, accessed 2003-01-06).

## 1.5 Spontaneous speech

The Lund intonation models are largely based upon studies of read, so-called 'laboratory' or 'lab speech'. The analysis of laboratory speech has allowed us to understand and model numerous prosodic phenomena. To a large extent, this understanding is of invaluable help as we move from the analysis of laboratory speech to the analysis of spontaneous speech. In general, the limited amount of control in spontaneous speech over different factors such as content word and syntactic structure makes analyses difficult. Patterns in spontaneous speech are more easily identified when we have a preconception about what we may find. Nevertheless, spontaneous speech has some unique features that are distinct from those of laboratory speech and, in some cases, it may be the case that the models we have developed for laboratory speech misdirect our attention. We may e.g. end up searching for categories in spontaneous speech which have been found in read speech as a result of our traditions of recitation, our conventions about prosodic patterns appropriate for so-called 'citation form' productions (see Beckman 1997 for a discussion).

In a typology of spontaneous speech, Beckman (1997: 7) defines 'spontaneous speech' as "speech that is not read to script". She furthermore distinguishes between ten different types of spontaneous speech recordings. The following is based on her typology. When making the decision to study spontaneous speech, one needs to

choose an elicitation technique that produces enough occurrences of the phenomenon one is interested in investigating, a recording sufficiently good for the planned analysis as well as a communicative situation allowing a reasonable degree of control over factors such as linguistic content and discourse structure.

The 'unstructured narrative' is elicited in an informal interview where the speaker is asked open-ended questions about e.g. his or her background. A skilled interviewer can elicit and record long monologue narratives with a high degree of audio quality using this technique. Most speakers seem to relax and forget that they are being recorded after a while, and therefore produce spontaneous speech with a high degree of naturalness. A disadvantage with the unstructured narrative is the lack of control over the content of the speech. An alternative to the unstructured narrative that involves a higher degree of control is the 'extended descriptive narrative'. It is obtained by asking subjects to retell a story. Prosodic phenomena in extended descriptive narratives have been studied in Swedish in e.g. Horne, Hansson, Bruce, Frid and Filipsson (2001). In 'instruction monologues', the speaker has been asked to instruct a real or imagined silent listener to perform a task. Good recordings and high control over both content words and syntactic structure can be obtained with this technique, unfortunately often at the expense of naturalness. The 'instruction dialogue' is comparable to the instruction monologue; it produces dialogues with a high degree of audio quality and control over the content of speech but somewhat unnatural speech. The 'database querying dialogue' technique resembles the instruction dialogue technique but produces a higher degree of naturalness as the speakers perform a task that they have initiated themselves (e.g. a railway timetable query). Database querying dialogues are, nevertheless, sometimes difficult to record as the speaker's consent must be obtained beforehand. An alternative then, is the 'Wizard of Oz' technique. The speaker uses a computer database querying system where the computer's responses are simulated, and performs a task that has been assigned to him or her. The Wizard of Oz technique suffers from the same disadvantage as the instruction monologue and dialogue technique, namely lack of naturalness, largely due to the fact that the task at hand is assigned to the speaker rather than initiated by the speaker him- or herself. Other types of spontaneous speech recordings discussed by Beckman are 'performance narrative' (e.g. recordings of after-dinner speeches), 'overheard conversation' (surreptitious recording of casual speech), 'enacted conversation' (conversation recorded from speakers who have given prior permission) and 'public conversation' (e.g. radio interviews).

Beckman exemplifies various areas of prosody research that she believes could benefit from studies of spontaneous speech. One of those areas is prosodic phrasing.

The speech material investigated in the present study comes from two databases. The study was undertaken within the HSFR/NUTEK financed research project *Swedish Dialogue Systems*, and the speech material collected for this project was employed in the first stage of the study (in chapter two). The investigated material from the *Swedish Dialogue System*'s speech database consists of dialogues between travel agents and clients. They were collected at travel bureaus in Lund, Skåne, from speakers who had given prior permission (see section 2.2.1 for further details). The dialogues are of the kind Beckman (1997) terms 'database querying dialogues' in the sense that they have a well-defined task-specified structure. However, moving from the study of the distribution of prosodic phrase boundaries to their phonetic realization, we felt it necessary to pay more attention to possible dialectal variation, and therefore use a speech material more carefully controlled for dialect. In chapters three to six, the speech material used is from the dialect project *SweDia 2000*'s database (see section 3.2.1 for further details). The investigated material is of the kind Beckman terms 'unstructured narrative'. It consists of informal interviews prompted with open-ended questions about e.g. the speakers' childhood or work. Thus, the speech material we have chosen to use in our investigations, at the expense of control, is spontaneous speech with a high degree of naturalness.

## 1.6 Aims of the study

This study deals with how prosody is used to divide the stream of speech into chunks indicating which words within a sentence belong together or form a syntactically, semantically and/or pragmatically coherent unit. These chunks of speech are referred to as 'prosodic phrases'. The primary aim of the study is to move away from the laboratory speech examined in many previous related studies and to investigate the phrasing function of prosody in spontaneous speech.

The problems to be dealt with in this study concern both the phonetics and the phonology of prosodic phrasing in spontaneous Swedish. As regards the phonetics of prosodic phrasing, we will examine how the prosodic variables duration (phrase-final lengthening), F0 and pausing are used to group words in speech. The combination of different variables, more specifically how they combine to signal boundary strength, will also be investigated. A phonological issue under investigation involves understanding what the production and perception constraints are that govern the grouping of words into prosodic phrases.

The planning of speech and the apparent potential in spontaneous speech for on-line changes in the speech plan are other issues that will be addressed, together with a discussion of prosodic phrase structure in spontaneous Swedish.

## 1.7 Outline

The present study comprises seven chapters. The introductory chapter is dedicated to defining what we mean by 'prosodic phrasing' and 'prosodic phrase', briefly reviewing the literature on prosodic phrasing in Swedish, and presenting the overarching goal of this study. The literature relevant for the specific research questions that are addressed in chapters two to six is reviewed in the chapters' introductory sections.

In the second chapter, we deal with a phonological issue. The distribution of prosodic phrase boundaries in spontaneous speech is investigated by considering it as a reflection of optimality theoretic constraints that restrain the production and perception of speech (e.g. constraints on the amount of speech that can be produced without making a planning stop and constraints on how the speech can be chunked up in relation to syntactic and information structure without rendering the speech incomprehensible to the listener). Based on a close review of the phrasing in a spontaneous dialogue material recorded at travel bureaus, a preliminary constraint hierarchy is proposed.

In chapter three, we move from the distribution of prosodic phrase boundaries to the phonetic realization of prosodic phrase boundaries. By examining articulation rate changes within the prosodic phrase, evidence of phrase-final lengthening, a reduction of the articulation rate in the final part of the prosodic phrase, is found. The frequent usage of phrase-final lengthening found in the data furthermore suggests that duration cues are just as important to prosodic phrase structure in spontaneous speech as they have been shown to be in read speech. The perceptual relevance of the unexpected traces of phrase-initial shortening (a higher articulation rate in phrase-initial than non-initial words) found in the data is discussed.

Chapter four presents an investigation of F0 downtrend in spontaneous Southern Swedish. We thereby shift our focus of attention from the phonetic signaling of phrase boundaries to the means used in speech to signal coherence within the prosodic phrase. By examining the relationships between F0 slope, phrase length and F0 starting point, we make an attempt to test the two Lund intonation models' capacities for describing spontaneous speech. The two approaches have different

implications for the amount of preplanning needed, which makes them particularly interesting to compare by testing spontaneous data.

Chapter five follows up on a finding made in chapter four, namely that F0 starting points vary, but seemingly not to accommodate for the length of the upcoming prosodic phrase. No feature of the prosodic phrase is found that accounts for the variation, and consequently, an explanation is sought in the discourse. More specifically, we investigate whether decreasing F0 starting points over the course of several prosodic phrases is used to signal coherence among phrases in spontaneous speech. Moving from investigations of the signaling of coherence within the prosodic phrase, we thereby turn to coherence signaling among prosodic phrases. The fact that neither does lowering of F0 starting points between prosodic phrases invariably give rise to the perception of coherence across the boundary (i.e. a weak boundary), nor does resetting invariably give rise to the perception of a strong boundary, suggests the presence of other cues that, in addition to intonation, affect the degree of perceived coherence.

Chapter six presents two perception experiments designed to relate perceived boundary strength to three known cues for prosodic phrasing in Swedish (pausing, F0 reset and final lengthening), and to investigate whether the established division into two phrasal categories or constituents (the 'prosodic phrase' and the 'prosodic utterance') has empirical support in spontaneous Swedish. In the introductory section, we furthermore discuss the 'intermediate phrase' as a possible prosodic constituent in Swedish. In chapter six, we follow up on the point made in chapter five that intonation alone does not explain the perception of boundary strength (or degree of coherence perceived between the phrases on either side of the boundary).

Chapter seven summarizes the main findings of the present study as well as the conclusions drawn.

CHAPTER 2

# Phrase boundary distribution

## 2.1 Introduction

In this chapter, we will present some hypotheses about optimality theoretic constraints on prosodic phrasing in spontaneous Swedish speech. The idea is to consider the distribution of prosodic phrase boundaries as a reflection of the constraints that restrain the production and perception of speech (e.g. constraints on the amount of speech that can be produced without making a planning stop and constraints on how the speech can be chunked up in relation to syntactic and information structure without rendering the speech incomprehensible to the listener).

Within 'Optimality theory', linguistic phenomena such as prosodic phrasing are described through the interaction of constraints (McCarthy and Prince 1993). One may assume that there are a number of constraints that interact to determine what prosodic form a sentence is assigned given its syntactic and information structure. Among these are constraints that both align prosodic phrase structure with syntactic and information structure as well as constraints that assign word and focal accents and delimit the prosodic phrases' size and accentual content. The constraints on

output representations are hypothesized to be universal. Languages differ only in the ranking order of the constraints.

### 2.1.1 Purpose

Since the constraints on output representations are hypothesized to be universal, it is interesting to investigate in what way constraints previously claimed for other languages can be assumed to interact in Swedish. The purpose of the present analysis of production data is to propose and discuss a possible constraint hierarchy for spontaneous Swedish. As a starting point, we take a number of constraints suggested to exist in other languages.

### 2.1.2 Optimality theory

Within optimality theory, linguistic phenomena are explained through the interaction of universal constraints. An optimal output form for a given input is selected from among a number of competing surface forms. The form that best satisfies the highest-ranking constraint, on which the candidates conflict, is considered to be optimal, as shown in Tableau 2.1. Constraints can be violated, but only minimally, i.e. only in order to satisfy a higher ranked constraint. The constraints are universal, but languages may differ with respect to each other in the ranking order of the constraints.

**Tableau 2.1** *An example of a constraint tableau*[1]*. Constraint A is ranked higher than constraint B (Constraint A >> Constraint B).*

| Candidates | Constraint A | Constraint B |
|---|---|---|
| ☞ $Cand_{1a}$ | | * |
| $Cand_2$ | *! | |

### 2.1.3 Previous studies

The constraints on prosodic phrasing have previously been discussed for a number of different languages, see e.g. Delais-Roussarie's (1996) constraints on French phrasing, Selkirk's (2000) constraints on English phrasing and Truckenbrodt's

[1] The optimal candidate is indicated with a '☞'. A candidate wrongly considered to be optimal is here indicated with a '✋'. The leftmost constraint in the tableau is the highest-ranking constraint and the rightmost constraint is the lowest-ranking. The cells of constraints that are irrelevant for the choice of optimal candidate are shaded. '*' marks a violation of a constraint and '*!' a fatal violation. The number of stars marks the number of violations.

(1999) constraints on phrasing in two Bantu languages (Kimatuumbi and Chichewa). Here, we take a constraint hierarchy suggested for English (Selkirk 2000) as a starting point in our discussion of the constraints on prosodic phrasing in spontaneous Swedish speech.

Selkirk (2000) has made some preliminary hypotheses about the hierarchy of constraints on prosodic phrasing in English. The constraints suggested to play a role are constraints on the syntax-phonology interface (Align-XP,R and Wrap-XP), constraints on the focus-phonology interface (Align-Focus,R) and constraints on the size and accentual content of prosodic phrases (Bin(MaP) and MiPAccent).

The constraints Wrap-XP and Align-XP,R are claimed to occupy the same rank in the English constraint hierarchy, see Tableau 2.2. Wrap-XP (Truckenbrodt 1995 and 1999) calls for the elements of an input morpho-syntactic constituent of type XP (maximal projection) to be contained within a prosodic constituent of type MaP (major phrase) in output representation. Align-XP,R, on the other hand, calls for the right (R) edge of any XP in syntactic structure to be aligned with the right edge of a phrase in prosodic structure. Wrap-XP comes into conflict with most (but not all) boundaries demanded by Align-XP,R.

**Tableau 2.2** *Wrap-XP, Align-XP,R (from Selkirk 2000: 242)*[2]

| [she [loaned] [her rollerblades]$_{NP}$ [to Robin]$_{PP}$ ]$_{VP}$ | Wrap-XP | Align-XP,R |
|---|---|---|
| ☞ a. (she loaned her rollerblades to Robin)$_{MaP}$ | | * |
| ☞ b. (she loaned her rollerblades)$_{MaP}$ (to Robin)$_{MaP}$ | * | |

A 'major phrase' is a prosodic domain above the level of the phonological word and the minor/accentual phrase in Selkirk's prosodic hierarchy. The major phrase corresponds to the 'intermediate phrase' in Pierrehumbert's (1980) work. As will be discussed in section 6.1.2, the intermediate phrase has no counterpart in the Swedish intonation model, and the next level of structure above the prosodic word (a domain with features of both the phonological word and the minor/accentual phrase, see section 6.1.2), is the 'prosodic phrase'. Thus, the relevant phrase in prosodic structure for the mapping between syntax and phonology, as well as information structure and phonology in Swedish is the prosodic phrase.

The Align-Focus,R constraint is ranked above the Wrap-XP and Align-XP,R constraints. Align-Focus,R calls for the right edge of a focus constituent in

[2] The broken line between the constraints indicates that they are not ranked with respect to each other.

information structure to be aligned with the right edge of a prosodic phrase. Ranked above the Align-Focus,R constraint, is the constraint MiPAccent[3]. MiPAccent calls for a minor phonological phrase to consist of at least one accent. Finally, below Wrap-XP and Align-XP,R in the constraint hierarchy, the constraint Binary(MaP), which calls for a major phrase to consist of just two minor phrases, is found.

In Selkirk's study, examples are given to support the hypothesis that the constraints Align-XP,R, Wrap-XP and Align-Focus,R, which interact with the phonological constraints on the prosodic phrases' size and tonal content, constitute a full account of English phrasing. However, she remarks that the results "need to be solidified on the basis of non-intuition-based investigation" (Selkirk 2000: 246).

## 2.2 Method

### 2.2.1 Speech material

Whereas most previous studies on constraints on prosodic phrasing have been limited to read speech, we have chosen to use spontaneous dialogues as data. Although we are aware of the fact that prosodic structure rarely is fully determinative, i.e. that several alternative prosodic realizations of a given syntactic structure are often possible, we assume that the prosodic phrasing produced by the recorded speakers is, in some sense, optimal.

The examined speech material consists of four spontaneous dialogues from the research project *Swedish Dialogue Systems'* database[4], recorded at travel agencies in Lund, Sweden. Four male and three female speakers' speech was analyzed[5]. The speech material was recorded on tape with a Sony WM-D6C stereo cassette recorder and two separate microphones, and subsequently digitized (with a sampling rate of 16 kHz and 16 bit resolution) and stored as ESPS/Waves+™ audio files.

---

[3] A DEPAccent constraint is also assumed, with the same ranking as MiPAccent, which assures that an accent in the output representation has a corresponding accent in the interface representation. Thus, the output is strictly dependent (DEP) on the input.

[4] Information about the HSFR/NUTEK financed research project *Swedish Dialogue Systems* (Grant No. F1472/1997) can be found at the following address: http://www.ida.liu.se/~nlplab/sds/ (Accessed 2003-01-06).

[5] One of the speakers appeared in more than one dialogue.

### 2.2.2 Prosodic transcription

The speech material was segmented and transcribed orthographically in the speech analysis program ESPS/Waves+™. After listening to the dialogues and visually inspecting F0 contours of the material, the transcription was completed with information about the distribution of prosodic phrase boundaries. Weak boundaries were indicated with '|' and strong boundaries with '||', in accordance with Strangert and Heldner's (1995a) proposal for boundary transcription within the base prosody system (Bruce 1994). Although a distinction was made between strong and weak boundaries in the transcription of the material, both boundary types have been dealt with in the same manner in the analysis below, since both mark the ends of prosodic phrases.

In two of the dialogues, the prosodic transcription was further completed with information about the positions of word accents (') and focal accents ('') which were also transcribed as in the base prosody system. Since word accent type (acute/accent I or grave/accent II) was of no relevance to this study, the symbol '`' was not used in combination with '' and '' to single out accented and focally accented accent II words from accent I words.

## 2.3 Empirical analysis

### 2.3.1 Speech repairs, Align-XP,R and Wrap-XP

Spontaneous speech is typically less structured syntactically than written discourse. It contains fragments and there is little subordination. A first observation about the prosodic phrasing in our data is that very few utterances have internal prosodic phrase boundaries, suggesting that Align-XP,R be ranked below Wrap-XP in spontaneous Swedish, see Tableau 2.3.

The pronoun *du* 'you' is a DP, i.e. a projection of a functional head D (determiner) with no complement (Abney 1987). According to the 'Lexical Category Condition' proposed by Truckenbrodt (1999), constraints that relate syntactic and prosodic categories apply to lexical syntactic elements and their projections, but not to functional elements and their projections[6], nor to empty syntactic elements and

[6] Lexical projections are projections of lexical heads (e.g. a N(oun) or V(erb)), whereas functional projections are projections of functional heads (Haegeman 1994).

their projections[7]. Thus, Align-XP,R does not call for an alignment of *du* with a prosodic phrase boundary. In contrast to pronouns, full NPs (which are also regarded as DPs by Abney (1987)) have a lexical NP complement. It is the lexical NP within the DP that can trigger a prosodic phrase boundary. DPs with a lexical NP complement are therefore marked as NPs in what follows.

du kan ändra datum på den |
$[[\text{du}]_{DP}\ [\text{kan ändra}\ [\text{datum}]_{NP}\ [\text{på den}]_{PP}\ ]_{VP}\ ]_{S}$
'you can change its date |'

**Tableau 2.3** *Wrap-XP >> Align-XP,R*

| $[[\text{du}]_{DP}\ [\text{kan ändra}\ [\text{datum}]_{NP}\ [\text{på den}]_{PP}\ ]_{VP}\ ]_{S}$ | Wrap-XP | Align-XP,R |
|---|---|---|
| ☞ $(\text{du kan ändra datum på den})_{PPh}$ | | * |
| $(\text{du kan ändra datum})_{PPh}\ (\text{på den})_{PPh}$ | *! | |

Utterances containing certain kinds of speech repairs constitute an exception to this ranking of the syntax-phonology constraints in the sense that these utterances sometimes violate the Wrap-XP constraint. In spontaneous speech, speakers do not produce perfect utterances. Speakers interrupt themselves, repeat and modify what they say. It has been shown that 'speech repairs' or 'disfluencies' are normal and frequent in spontaneous speech (Heeman 1997, Nakatani and Hirschberg 1994, Schriberg 1994). What is the optimal way of phrasing an utterance that contains a speech repair? How does the speaker let the hearer know that a part or all of what has just been said should be disregarded? Although disfluencies are not seen as a part of the grammar, they are produced and the prosodic strategies used to repair them (make the hearer aware of them and signal how to interpret them) have to be considered an important part of a speaker's prosodic phonology.

Heeman has shown for English that the end of a disfluency reparandum (the interruption point) is often accompanied by a disruption in the intonation contour. This indicates that speech repair phenomena need to be taken into consideration when formulating the constraints on prosodic phrasing in spontaneous speech.

Heeman divides speech repairs into three different categories: 'fresh starts', 'modification repairs' and 'abridged repairs'. 'Fresh starts' occur where the speaker abandons the current utterance and starts again. Heeman claims that "they are defined in terms of a strong acoustic signal marking the abandonment of the

[7] A nonexhaustive phrasing where only the VP would be contained within a prosodic phrase is excluded by a requirement of the parsing to be exhaustive on every prosodic level.

current utterance" (1997: 12) and that the interruption point "often is accompanied by a disruption in the intonational contour" (1997: 9). Intonational disruptions are also found in fresh starts in Swedish, see e.g. Figure 2.1. Therefore, we may choose to allow the syntax-phonology constraints to also call for the interruption points of abandoned, incomplete XPs to be aligned with prosodic phrase boundaries, at least the interruption points of fresh starts.

om jag titt(ar) | fjortonde har jag plats ||

reparandum interruption point

'if I loo(k) | on the fourteenth I have seats ||'

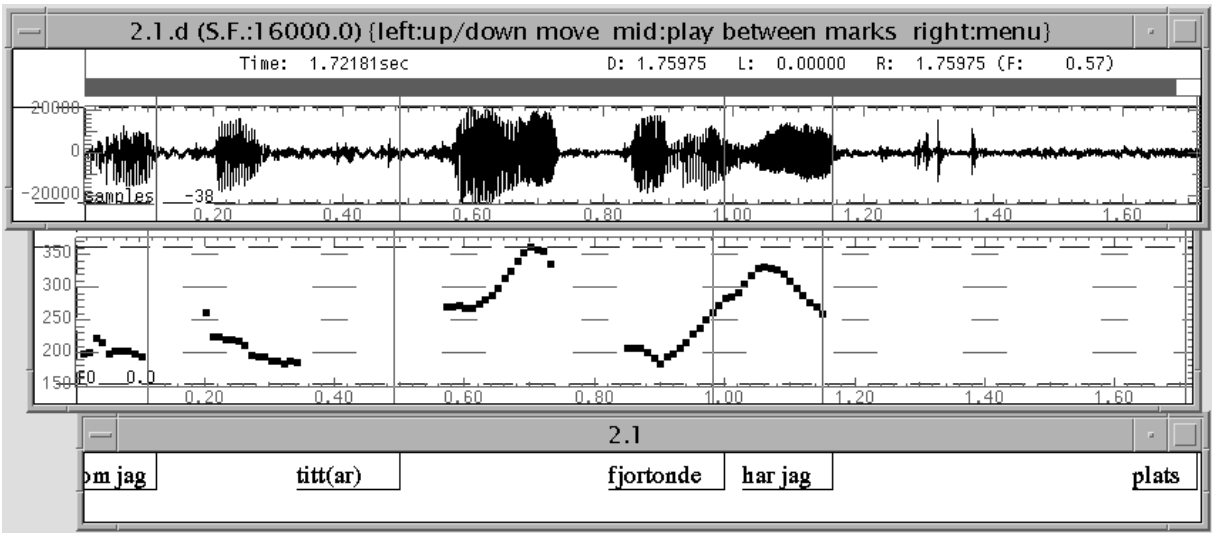


**Figure 2.1** *Speech wave and F0 contour of* om jag titt(ar) fjortonde har jag plats *'if I loo(k) on the fourteenth I have seats' (female speaker of Southern Swedish).*

The 'modification repairs' have a strong word correspondence between reparandum and alternation, see e.g. Figure 2.2. Heeman suggests that the word correspondence can help the hearer to determine the extent of the reparandum as well as function to signal that a modification repair has occurred. Therefore, an acoustic marking of the interruption point of a modification repair, e.g. a boundary tone or a F0 reset, does not appear to be necessary, as confirmed by the examined examples here.

'Abridged repairs' consist of an editing term, e.g. a hesitation sound/filled pause, but no reparandum (Heeman 1997), see e.g. Figure 2.3. In spontaneous speech, filled pauses are frequent e.g. after conjunctions and before important content words. However, as the phrase is continued after the repair, we do not expect the Align-XP,R constraint to call for an alignment of the abridged repair with a prosodic phrase boundary. A cohesional strategy is more common.

vi har ju Air F vi har ju SAS också

*reparandum* *alternation*

'we have Air F we have SAS as well'

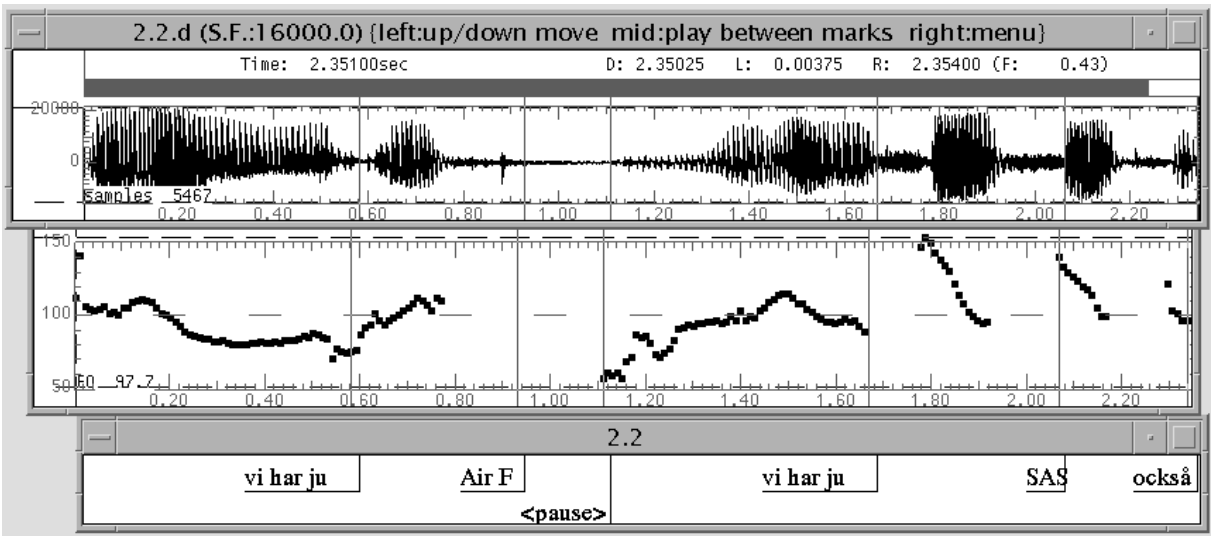


**Figure 2.2** *Speech wave and F0 contour of* vi har ju Air F vi har ju SAS också *'we have Air F we have SAS as well' (male speaker of Stockholm Swedish).*

nej för att jag hade bytt äh datum nämligen ||

*editing term*

'because I had changed uh the date you see ||'

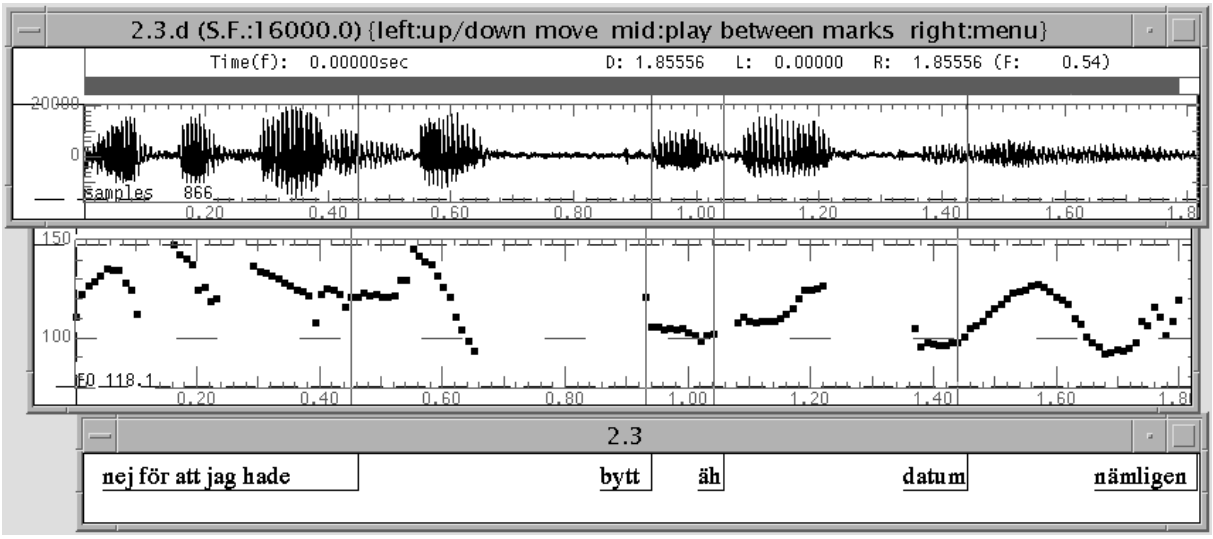


**Figure 2.3** *Speech wave and F0 contour of* nej för att jag hade bytt äh datum nämligen *'because I had changed uh the date you see' (male speaker of Southern Swedish).*

In summary, it is necessary to allow the syntax-phonology constraints to also apply to abandoned XPs in a constraint hierarchy for prosodic phrasing in spontaneous speech. Whereas abridged repairs do not appear to need an acoustic marking of their interruption point, fresh starts do. Modification repairs are not necessarily acoustically marked since they appear to rely on a strong word correspondence to delimit the extent of the reparandum.

When examining the prosodic correlates of speech repairs, it appears that it may be necessary to divide the Align-XP,R constraint into two distinct constraints: Align-XP,R and Align-XP,L. The Align-XP,R constraint would be responsible for associating the right edges of a XP with boundary signaling cues (e.g. a boundary tone and final lengthening) and the Align-XP,L constraint for associating the left edge of a XP with boundary cues (e.g. a resetting of F0). One would then assume that the Align-XP,L constraint applies to each XP or abandoned XP, while the Align-XP,R only applies to complete XPs. Such an assumption can be made on the basis of the observation that the interruption points of speech repairs do not always show final lengthening and a boundary tone.

### 2.3.2 Root sentences, Align-XP,R and Wrap-XP

Certain types of constructions, e.g. parenthetical expressions, introduce internal prosodic phrase boundaries by forming intonation domains on their own within the root sentence. These constructions are external to the root sentence with which they are associated (Nespor and Vogel 1986). In (2a) the parenthetical, clarifying expression, a so-called right detachment or antitopic construction (Lambrecht 1994), is found at the end of the root sentence. Here we may assume that Wrap-XP wraps the root sentence and the external construction (E.C.) separately, as two distinct morpho-syntactic constituents.

(2a)
nej det är den billigaste jag har | Air France där ||
[nej det är den billigaste jag har]$_{ROOT}$[Air France där]$_{E.C.}$
'no it's the cheapest I have | Air France there ||'

The phrasing in Figure 2.4 can be described in the same manner, i.e. as the result of Wrap-XP wrapping the root sentence and the clarification separately, as two distinct morpho-syntactic constituents.

och det är bara här jag kan göra det då | köpa den hos er ||
[och det är bara här jag kan göra det då]$_{ROOT}$ [köpa den hos er]$_{E.C.}$
'and it's only here I can do that then | buy it here from you ||'

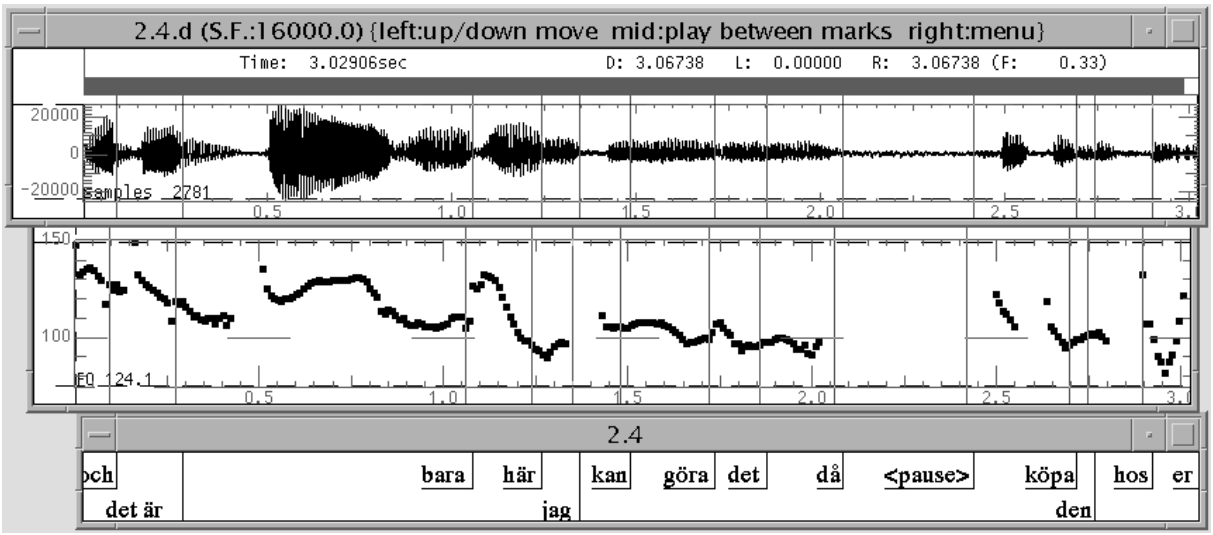


**Figure 2.4** *Speech wave and F0 contour of* och det är bara här jag kan göra det då köpa den hos er *'and it's only here I can do that then buy it here from you' (male speaker of Southern Swedish).*

However, in (2b) the external construction is internal to the root sentence. A violation of Wrap-XP would appear to be inevitable. If the external construction itself is wrapped, then the larger XP is not, i.e. the larger XP is not contained in a single prosodic phrase and vice versa. With the violation of Wrap-XP, Align-XP,R right-aligns all XPs with the edges of prosodic phrases except the noun phrase *ett sådant kort* 'one of those cards' which will be discussed below in section 2.3.3.

(2b)
jo du måste ha ett sådant kort till | ett sådant | till bägge de här va ||
[jo du måste ha ett sådant kort till [ett sådant]$_{E.C.}$ till bägge de här va]$_{ROOT}$
'yes you have to have one of those cards for | one of these | for both of these right ||'

## 2.3.3 Constraints on accentual content and size

The utterances in the examined domain, i.e. the travel domain, are, in general short and they are usually wrapped into a single prosodic phrase. This tendency provides support to a higher ranking of the Wrap-XP constraint than the Align-XP,R constraint. However, some evidence to support the hypothesis of an interaction between the syntactic constraints and constraints on the size and tonal content of prosodic phrases has been found.

Since the defining feature of a prosodic word is that it contains an accent (see section 1.4), we may formulate a constraint PWdAccent that sets the minimal tonal content of a prosodic word to one accent. Because a prosodic phrase consists of at least one prosodic word, we can formulate a constraint Min(PPh) that demands that a prosodic phrase contain at least one accent. Min(PPh) can be assumed to be the constraint which prevents *till* 'to/for' from constituting a prosodic phrase on its own (i.e. the phrasing in candidate *f* in Tableau 2.4). In order to do so, the Min(PPh) constraint must be ranked above Wrap-XP and Align-XP,R (i.e. Min(PPh) >> Wrap-XP >> Align-XP,R).

Below, we treat the incomplete preposition phrase *till* 'to/for' as if it were a complete phrase based on the reasons given in section 2.3.1 above, i.e. we expect Align-XP,R to require that its right edge be aligned with a prosodic phrase boundary. The unaccented word *va* 'right' does not constitute a prosodic word on its own and has therefore been included in the last prepositional phrase.

jo du ''måste 'ha ett 'sådant ''kort till | ett 'sådant | till ''bägge de 'här va ||
[jo du [måste ha [ett sådant kort]$_{NP}$ [till]$_{PP}$ [ett sådant]$_{NP}$ [till bägge de här va]$_{PP}$ ]$_{VP}$ ]$_{S}$
'yes you have to have one of those cards for | one of these | for both of these right ||'

**Tableau 2.4** *Min(PPh) >> Wrap-XP >> Align-XP,R*

| [jo du [måste ha [ett sådant kort]$_{NP}$ [till]$_{PP}$ [ett sådant]$_{NP}$ [till bägge de här va]$_{PP}$ ]$_{VP}$ ]$_{S}$ | Min (PPh) | Wrap-XP | Align-XP,R |
|---|---|---|---|
| a. (jo du måste ha ett sådant kort till ett sådant till bägge de här va)$_{PPh}$ | | * | *!** |
| b. (jo du måste ha ett sådant kort)$_{PPh}$ (till ett sådant till bägge de här va)$_{PPh}$ | | * | *!* |
| c. (jo du måste ha ett sådant kort till)$_{PPh}$ (ett sådant till bägge de här va)$_{PPh}$ | | * | *!* |
| d. (jo du måste ha ett sådant kort till ett sådant)$_{PPh}$ (till bägge de här va)$_{PPh}$ | | * | *!* |
| ☞ e. (jo du måste ha ett sådant kort till)$_{PPh}$ (ett sådant)$_{PPh}$(till bägge de här va)$_{PPh}$ | | * | * |
| f. (jo du måste ha ett sådant kort)$_{PPh}$ (till)$_{PPh}$ (ett sådant)$_{PPh}$ (till bägge de här va)$_{PPh}$ | *! | * | |

The extent of a prosodic phrase can be described in a number of ways. We can e.g. characterize a prosodic phrase as containing a certain number of syllables or feet, we can measure its duration in time or count the number of focal and/or non-focal accents it contains.

In Horne and Filipsson (1998), it is suggested that syllable count plays a role in determining the position of prosodic phrase boundaries. In their material (read speech), where the speech rate was on the average of about 5 syllables per second, prosodic phrases contained between 7 and 63 syllables, with a mean at 24 syllables (s.d.=10.3 syllables). In the spontaneous (unscripted) speech examined in the present study, the prosodic phrases contain a much smaller number of syllables.

However, because there are few long syntactic structures in the examined material, we are not, at this point, able to determine if the observed units are maximal optimal chunks for linguistic processing in spontaneous speech, or if Wrap-XP would wrap even larger constituents. Examples such as (2c) and (2d) indicate that the syntax-phonology constraints' force may be delimited by maximal and minimal size constraints, and that these constraints not only have the ability to divide the larger maximal projection VP into several prosodic phrases, but also to group functional projections. According to Truckenbrodt (1999), Wrap-XP only demands that lexical projections are grouped together, not functional projections. In other words, the phrasing in (2c) would not violate Wrap-XP even if *kronor* 'crowns' had been right-aligned with a prosodic phrase boundary. This indicates the existence of a cohesional force that is stronger than that which is demanded by Wrap-XP.

(2c)
det kostar nittio kronor om du inte har det |
[det [kostar [nittio kronor]$_{NP}$ ]$_{VP}$ [[om]$_{COMP}$ [du [inte har det]$_{VP}$ ]$_{S}$ ]$_{S'}$ ]$_{S}$
'it's ninety crowns if you don't have one |'

Possibly it is the unit XP to which the phrase in the prosodic structure is mapped that is problematic in our analysis. The morpho-syntactic unit XP allows for consistent identification, but we know that the prosodic phrasing structure, despite its close relation to syntax, is not strictly dependent of syntactic structure. There is no isomorphic relationship between the two levels or hierarchies (see section 1.1). This may indicate that the mapping between levels does not occur between syntax and phonology, but between some higher level in the speech planning and phonology. It would also explain the "discrepancy between the syntactically motivated surface structure and what is apparently required as an input to the phonological component" reported by Chomsky and Halle (1968: 372).

In (2d), not only the edges of maximal projections are right-aligned with prosodic phrase boundaries (as demanded by Align-XP,R), but also the verb *undrar om* 'wonder about'. This tendency for inserting a pause and/or a weak phrase boundary

before important content words or phrases, has previously been reported by Strangert (1993), and appears to be a characteristic of spontaneous speech. It might be analyzed as a reflection of an Align-Focus,L constraint. Information structure-phonology constraints will be discussed in further detail in the next section.

(2d)
jag undrar om | en resa till Phuket | den tolfte trettonde december ||
[jag [[undrar om]$_{V}$ [en resa [till Phuket]$_{PP}$ [den tolfte trettonde december]$_{NP}$ ]$_{NP}$ ]$_{VP}$ ]$_{S}$
'I'm wondering about | a trip to Phuket | on the twelfth thirteenth of December ||'

In summary, it is necessary to assume that there is an interaction between the syntactic constraints and further constraints on the maximum and minimum size of prosodic phrases. However, the examined material does not allow us to draw any conclusions as to their nature with the exception of Min(PPh) that sets the minimal tonal content of a prosodic phrase to one accent. Finally, evidence to support the existence of a cohesional force stronger than the Wrap-XP has been found.

### 2.3.4 Maximal focus projection and Align-Focus,R

In English, the constraint Align-Focus,R, ranked above Wrap-XP and Align-XP,R aligns the right edge of a focus constituent in information structure with the right edge of a prosodic phrase (Selkirk 2000).

The focus, the part of the sentence that conveys new information, may cover a larger or smaller part of the sentence (see section 1.4). In a sentence with a 'broad focus', the focus may cover the entire sentence (a maximal focus projection). According to the 'Focus-to-Accent' (FTA) approach, the focus of a sentence is marked by a focal accent (Gussenhoven 1983, Ladd 1996). If the Align-Focus,R constraint is satisfied, the focus constituent's right edge is also aligned with a prosodic phrase edge.

When examining utterances with foci covering the entire utterance and foci in phrase-final position, a high ranking of the Align-Focus,R constraint also seems to be relevant for Swedish, see Figure 2.5 and Tableau 2.5. This structure does not lead to a conflict between Wrap-XP and Align-Focus,R. Here a single prosodic phrase containing the entire structure allows both the focus constituent to be right-aligned with a prosodic phrase and XP to be contained within a single prosodic phrase.

jag ˈglömde ˈplånboken ˈhemma |
[jag glömde plånboken hemma]$_{Foc}$
'I forgot my wallet at home |'

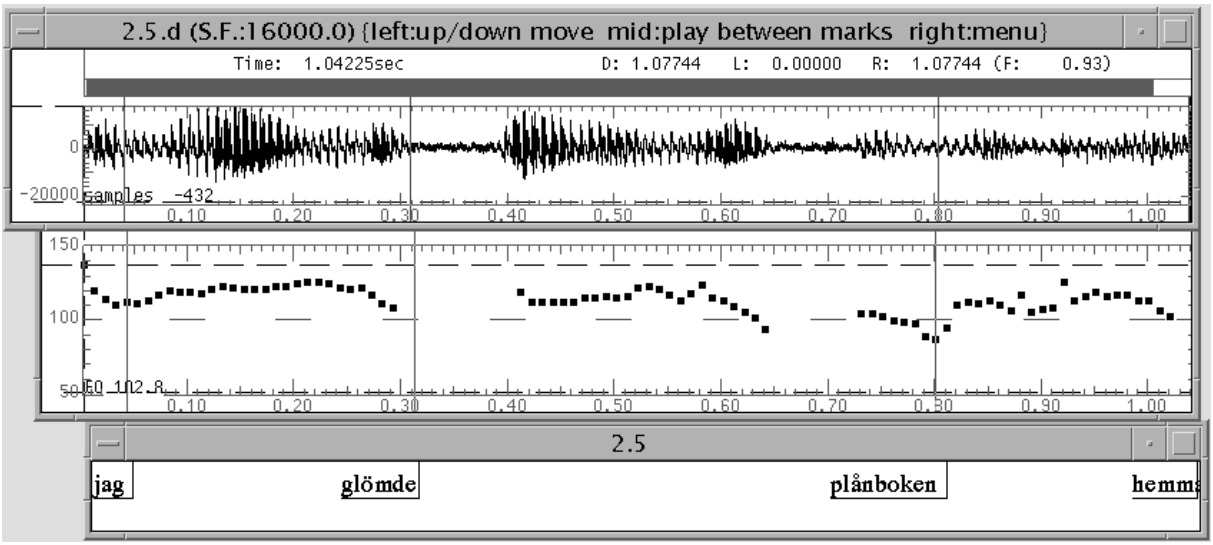


**Figure 2.5** *Speech wave and F0 contour of* jag glömde plånboken hemma *'I forgot my wallet at home' (male speaker of Southern Swedish). The lack of a focal accent marking the focus is not unusual in Southern Swedish.*

**Tableau 2.5** *Align-Focus,R >> Wrap-XP >> Align-XP,R*

| [jag [glömde [plånboken]$_{NP}$ hemma]$_{VP}$ ]$_{S\text{-}Foc}$ | Align-Focus,R | Wrap-XP | Align-XP,R |
|---|---|---|---|
| ☞ a. (jag glömde plånboken hemma)$_{PPh}$ | | | * |
| b. (jag glömde plånboken)$_{PPh}$ (hemma)$_{PPh}$ | | *! | |

However, when focus is in non-final position, it is apparent from our data that the right edge of a focus constituent is not necessarily aligned with the right edge of a prosodic phrase in spontaneous Swedish speech, see Figure 2.6. In most dialects of Swedish[8], e.g. in Stockholm Swedish, a phrase accent is found after the word accent fall in words in focal position. In the English major phrase or intermediate phrase, the phrase accent functions as an edge tone. However, as there is no obligatory deaccentuation of lexical items postfocally in Stockholm Swedish, the phrase accent is not an edge tone. The phrase accent is not peripheral to a phrasal constituent, which is why the Align-Focus,R constraint is not relevant for Swedish. The issue of the phrase accent and the intermediate phrase will be discussed in further detail in section 6.1.2.

[8] In Southern Swedish dialects, the phrase accent can be described as a L tone. However, since the word accents may be described as falling HL accents (see section 4.4.1.2), the phrase accent is not clearly observable or distinguishable from the word accent L, and the focal accents can therefore be described as 'boosted' HL accents rather than HLL$^{-}$ accents.

[vilken ''månad]$_{Foc}$ ska du 'åka i |
[vilken månad]$_{Foc}$ ska du åka i
'what month are you leaving in |'

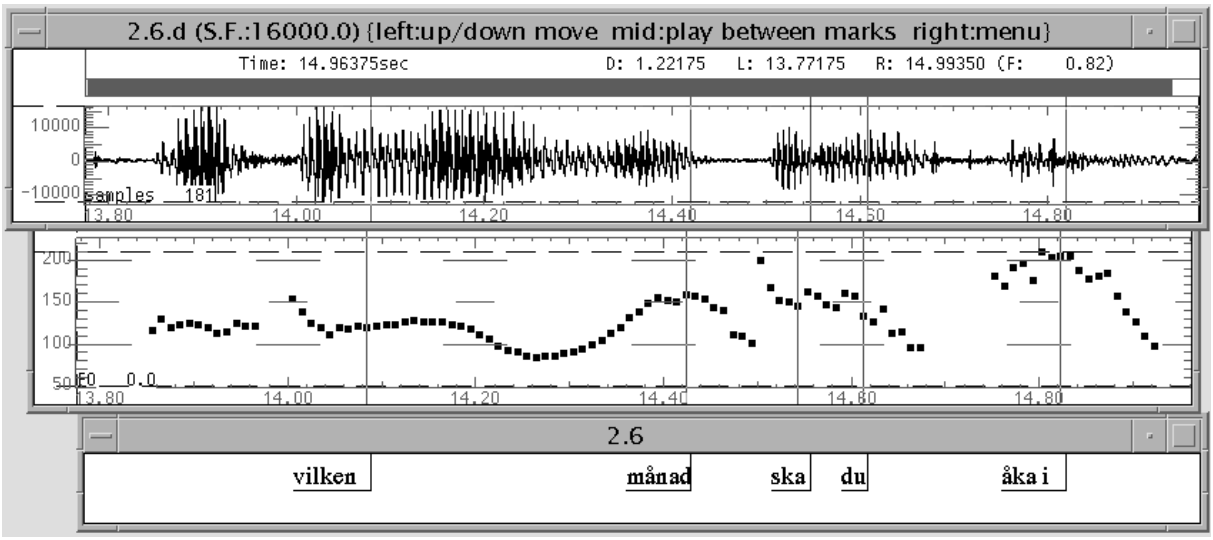


**Figure 2.6** *Speech wave and F0 contour of* vilken månad ska du åka i *'what month are you leaving in' (male speaker of Stockholm Swedish).*

In Horne, Hansson, Bruce and Frid (2001), a late timing of the phrase accent (a H tone) was found in focally accented utterance-initial words in Stockholm Swedish. The late timing of the H$^-$ was found both on words in utterance-initial foci and on so-called 'contrastive topics' (Hansson 2001) and it was analyzed as a coherence cue.

In summary, the right edge of a focus constituent is not necessarily aligned with the right edge of a prosodic phrase in spontaneous Swedish speech. Quite the contrary, in terms of prosodic phrasing, it has been suggested that the observed 'delay' of the H$^-$ sometimes found in Stockholm Swedish can be interpreted as a way of explicitly making the focally accented word a part of the same prosodic phrase as the information that follows.

## 2.4 Summary

In the present chapter, an attempt has been made to determine whether empirical evidence can be found for theoretical claims made about a number of universal constraints on prosodic phrasing. Based on analyses of production data, a possible constraint hierarchy for spontaneous Swedish has been proposed and discussed.

As pointed out by Selkirk (2000), as far as the syntax-phonology interface is concerned, languages must opt for a dominant cohesional strategy or a dominant

demarcative structure, as represented by the relative rankings of the constraints Wrap-XP and Align-XP,R. In the case of Swedish, the data analyzed indicate a cohesional strategy. However, the speech style influences the choice of phrasing strategy. Both rate and style of speech have an effect on the length of the prosodic phrases. Therefore, it is plausible to assume that the constraint ranking described in the present chapter is a ranking specific for spontaneous Swedish.

Examination of so-called disfluencies or speech repairs revealed two different strategies: a cohesional strategy across 'abridged repairs' and 'modification repairs', and a demarcative strategy helping listeners to identify 'fresh starts'. Differences in the realization between phrase boundaries at fresh starts and other positions, suggested a division of the Align-XP constraint into two: Align-XP,L and Align-XP,R.

Although further constraints on prosodic phrases' maximum and minimal size need to be added to the hierarchy, the following preliminary ranking for Swedish spontaneous speech was proposed: Min(PPh) >> Wrap-XP >> Align-XP,R. Evidence against a high ranking of the constraint Align-Focus,R in Swedish was found which can be related to the non-peripheral nature of phrase accents in Swedish. However, a strategy by which the prosodic marking of the focal information is reinforced by left-aligning it with a prosodic phrase boundary was observed, i.e. evidence of an Align-Focus,L constraint.

CHAPTER 3

# Articulation rate in boundary signaling

## 3.1 Introduction

Moving away from the phonology of prosodic phrasing in spontaneous speech, this chapter presents results from a study on the phonetic signaling of boundaries in spontaneous speech. Previous studies on the production of prosodic phrase boundaries have shown that a whole constellation of cues are involved in boundary signaling in read speech. Taking these studies as a starting point, we will investigate some of the prosodic means used to group words and chunk speech in spontaneous speech. Moving from the distribution to the phonetic realization of prosodic phrase boundaries, we need to pay more attention to possible dialectal variation, and therefore use a speech material that is more carefully controlled for dialect. In the following, all speech material used is representative for the dialect spoken in *Skåne*. In the present chapter, we investigate phrase-final lengthening as a cue to prosodic phrase structure, whereas F0 reset will be examined to some extent in chapters four, five and six and pausing in chapter six.

### 3.1.1 Phrase-final lengthening in Swedish

In the present chapter, we look for evidence of a reduction of the articulation rate within the prosodic phrase that could be interpreted as 'phrase-final/final lengthening' or 'constituent-final lengthening'[1]. We follow Lindblom (1978: 85) in taking final lengthening to mean the effect caused to a unit of spoken language (or music) that "has a longer duration when, within a larger unit, it occurs finally".

Final lengthening is clearly visible in Swedish read production data (Bruce *et al.* 1991, Horne, Strangert and Heldner 1995), and has also convincingly been shown to be an important cue in the perception of phrasing in Swedish (Bruce *et al.* 1993). In the series of perception experiments undertaken in Bruce *et al.* (1993), it was shown that increased duration in segments is perceived as a cue for a boundary whereas reduced duration signals coherence. Both the boundary and coherence signaling cues were important for the listeners' perception of the boundary location in a syntactically ambiguous test sentence. Although most listeners had relied on a combination of duration and F0 cues for their decisions, some listeners primarily paid attention to the duration cues. Furthermore, some evidence was presented to suggest that listeners who perceive segment durations as the primary correlate to prosodic phrase structure also largely rely upon duration cues in their own production.

### 3.1.2 Interpretations of final lengthening

Lindblom (1978) regards the different interpretations of final lengthening found in the literature as explanations along three major avenues: 1) Final-lengthening is a learned and language-specific phenomenon, 2) final lengthening signals constituent structure to the listener (i.e. is perceptually motivated), and 3) final lengthening is a consequence of speech production constraints (the short term memory model (Cooper 1976, Lindblom, Lyberg and Holmgren 1976), the power law model (Lehiste 1970, Lindblom *et al.* 1976), the F0-dependence model (Klatt 1975, Lyberg 1977, 1978 and 1981) and the planning vs execution model (Cooper 1976)). The short term memory model relates final lengthening to the capacity of the memory component, the power law model is a derivative of the constant word duration model, and the planning vs execution model views final lengthening as

[1] A related term is 'prepausal lengthening'. However, since the phenomenon in question also occurs in prosodic phrases that are not separated from the following phrase by a pause (silent interval), this term is not suitable to use here.

reflecting a slowing down of execution during the planning of subsequent gestures. The F0 dependence model will be discussed in further detail in section 3.1.3.

Four facts are presented and discussed which lead Lindblom to conclude that there is no direct phonetic causation of the final lengthening phenomenon: 1) the degree of final lengthening varies across languages, 2) final lengthening is less marked in infant babbling than in adult speech, 3) final lengthening is not present in the speech of deaf speakers, and 4) final lengthening occurs not only before pauses but also before syntactic boundaries. Lindblom therefore concludes that final lengthening is a learned and language-specific phenomenon but phonetically natural. The latter part of his conclusion rests on the observation that phonetically natural processes are easier to learn and to use and therefore stand a better chance of being incorporated into a language's phonology. Lindblom's interpretation of final lengthening thereby explains why it has emerged in so many languages without assuming the existence of a direct link between phonetic cause (e.g. F0 dependence) and speech behavior.

Above, we have chosen to elaborate mostly on the view of Lindblom (1978). In addition to the arguments Lindblom presents to show that final lengthening is learned and language-specific, there is evidence in the literature suggesting an opposite dependency relation between F0 and duration to that which is assumed in the F0 dependency model, i.e. an accommodation of F0 to the amount of sonorant segmental material available. Grabe (1998), among others, has shown that there are two ways in which languages accommodate accents when sonorant segmental material is scarce. In (at least some dialects of) English, accents are compressed (reflected in an increase in the rate of F0 change) whereas in Northern Standard German, accents are truncated (reflected in F0 movements that end early). We know that these strategies are employed in Swedish as well, although possibly in different ways in different dialects (Alstermark and Erikson 1971, Bannert 1982, Bannert and Bredvad-Jensen 1975).

### 3.1.3 Dialectal variation

In Lyberg (1977, 1978 and 1981), the phrase-final lengthening phenomenon in Swedish is related to the distribution of prominence within the phrase, more specifically to the phrase-final focal accent (the so-called 'default sentence accent').

In Swedish prosody research, although a great deal of attention has been paid to dialectal variation in word and phrase accent realization (see e.g. Bruce and Gårding 1978, Bruce 2001), possible differences as regards the distribution of accents within

the phrase have not been investigated in many studies. Both the realization and distribution of accents are, however, relevant for our understanding of the internal structure of the prosodic phrase.

A sketch of dialectal differences in prominence distribution within the phrase has, nevertheless, been made by Gårding (1974) and Gårding, Bannert, Bredvad-Jensen, Bruce and Nauclér (1974). Gårding (1974) notes that the most prominent accent, the focal accent or the 'central stress' (*centralbetoningen*) in her terminology, is usually found at the end of the phrase in both Southern Swedish (*skånska*) and Stockholm Swedish. She relates this observation to the fact that the highest informational load is often found in the predicate (the verb phrase) which in Swedish is generally found after the subject, i.e. in the later part of the sentence. The late placement of the focal accent can be related to principles such as the 'Given before New Principle', the 'Theme First Principle' and the 'Discourse Iconicity Principle'. All principles place 'given' or 'thematic' information early in the sentence, close to the part of the discourse to which it relates, and 'new' or 'focal' information (the 'rheme', the information with the highest informational load) in the later part of the sentence[2], after a frame has been given to direct the listener's attention (for an overview of the relevant literature, see Lambrecht 1994). Nevertheless, in phrases classified by Gårding as having semantically equally heavy elements, different rules apply in Swedish dialects. Stockholm Swedish is labeled right-dominated (stress falls on the last element), whereas Southern Swedish is observed to have a more equal spread of prominence in the phrase or to be left-dominated, see Table 3.1. That a central stress or default sentence accent cannot always be identified in Southern Swedish phrases has also been reported by Grønnum Thorsen (1988), and our own observations indicate that equal spread of prominence is found in other contexts than those mentioned by Gårding (1974), e.g. in broad foci (see section 2.3.4).

[2] A discussion of the different definitions used in the literature for the terms given above can be found in Molnár (1998).

Table 3.1 *Examples of the prominence distribution in noun phrases in Southern and Stockholm Swedish (from Gårding 1974: 55). A double bar indicates so-called central stress, single bars indicate an even prominence distribution.*

| Southern Swedish (*skånska*) | Stockholm Swedish | English translation |
|---|---|---|
| ˈfem ˈår *(slow speech tempo)*<br>ˈˈfem år *(fast speech tempo)* | fem ˈˈår | 'five years' |
| ˈfem ˈår gammal *(slow)*<br>ˈˈfem år gammal *(fast)* | fem år ˈˈgammal | 'five years old' |

Finally, in addition to discussing the distribution of prominence in the phrase, Gårding (1974) also gives an account of some dialectal differences concerning the phonetic realization of prominence. Evidence that the phonological and/or phonetic conventions of a standard variety are not always applicable to all varieties of a language has also been found in other, more recent studies of dialectal intonation variation in English (see e.g. Fletcher, Grabe and Warren forthc).

There are several differences in the accent realization and accent distribution between standard and Southern Swedish that may have implications for the phrasing strategies used. For example, Southern Swedish shares many prosodic properties with Danish (Gårding *et al.* 1974), a language that has been claimed to lack phrase-final lengthening in some dialects (Grønnum Thorsen 1988). This is a fact that supports the claim that final lengthening is a learned and language-specific phenomenon.

Like Danish (Grønnum Thorsen 1988), Southern Swedish lacks a (high) phrase accent, the high turning point that in Stockholm Swedish follows the word accent fall in focal and phrase-final positions (see section 1.4). Lyberg (1977, 1978 and 1981) has suggested that the phrase-final lengthening phenomenon in Swedish is related to this characteristic phrase end contour of the fundamental frequency, the rise-fall gesture after the last word accent fall. In Southern Swedish, however, there is no such gesture after the last word accent and there would therefore not be a need for prolonging the phrase-final durations or slowing down the articulation rate phrase-finally. Furthermore, as in Danish, there is no so-called default sentence accent or focal accent at the end of the phrase in Southern Swedish. In other words, the most prominent accent is not always found on the last word of the phrase. In many cases, there is no accent that is clearly assigned more prominence, as discussed in the previous section. Thus, even assuming that focal accentuation in itself (regardless of the tonal gesture associated with it) results in final lengthening, there

is no reason to believe that final lengthening is obligatory in the Southern Swedish phrase.

On the other hand, it can be argued that, despite phonological differences, the reflection of the tones as observed in the F0 contour at the end of the prosodic phrase differs little between Stockholm and Southern Swedish, see Figure 3.1. In the word investigated by Lyberg (1981), *dag* 'day', the F0 contours are almost identical in the two dialects. First, F0 rises to the H in the word accent H*L in Southern Swedish, and from the L* in the word accent (H)L*[3] to the phrase accent H⁻ in Stockholm Swedish. Then, F0 falls from the H to the L in the word accent H*L in Southern Swedish (and to a possible boundary tone, L%, which cannot be clearly distinguished from the L in the H*L accent[4]), and from the phrase accent to the boundary tone L% in Stockholm Swedish. In the light of this, the prerequisites for final lengthening's F0 dependence can be argued to exist in both dialects.

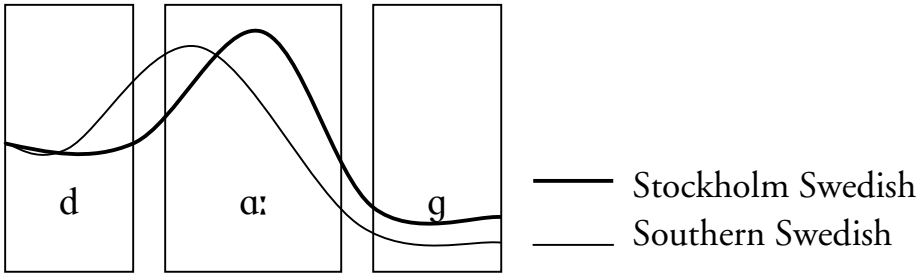


**Figure 3.1** *Stylized F0 contours of the focally accented word dag 'day' phrase-finally in Stockholm Swedish, (H) L*H⁻ L%, and Southern Swedish, H*L (L%). Based on Bruce (1977: 45) and Bruce and Gårding (1978).*

Regardless of whether final lengthening in Southern Swedish is F0 dependent or not, it is interesting to look for non-tonal cues for prosodic phrasing in this dialect. The prosodic phrase in Southern Swedish does not end with a prominent rise-fall gesture (after the last word accent), but rather with an accent that does not deviate from preceding ones, neither as regards direction nor necessarily the extent of its F0 movement (much like Copenhagen Danish). Further, since the word accents in Southern Swedish end low, the L boundary tones are less perceptually salient than in Stockholm Swedish (where the L boundary tone follows a H phrase accent). Therefore, non-tonal cues may prove very important in Southern Swedish.

[3] In one-syllable accent I-words, the H in the HL* accent is not realized.

[4] A L phrase accent, L⁻, is assumed between the word accent L and the L boundary tone in some phonological analyses of Southern Swedish. We choose not to do so in our analysis of Southern Swedish based on the observation that increasing prominence both raises the H and lowers the L of the word accent.

### 3.1.4 Articulation rate variation

A method for measuring final lengthening is developed and discussed in Wightman *et al.* (1992). Since the inherent duration of a phone is known to be the largest source of variation in segmental duration, in measuring final lengthening one would like to measure the difference between the duration of a given segment and the mean duration of that specific phone type. Wightman *et al.* (1992) therefore introduce 'normalized duration', which measures the duration of a segment as the number of standard deviations from the mean duration of the phone contained in the segment. In order to obtain the means and standard deviations of a speaker's phones, a fairly large amount of speech data needs to be segmented and labeled. This can be done automatically, as has been done for Stockholm Swedish (Lindström, Bretan and Ljungqvist 1996), but since we have no speech recognizer available to us for segmenting and labeling our Southern Swedish data, we will choose another method in the present investigation.

In an experiment conducted to determine if articulation rate variation has a specific domain in spontaneous speech (Czech), Dankovičová (1997) found a regular pattern within the intonational phrase. By measuring and comparing the articulation rate in each phonological word in the phrase, she was able to show that the articulation rate slowed down over the course of the phrase. The first or second word was demonstrated to have the highest articulation rate and the last word to have the slowest. The so-called 'interpause stretch' (stretch of speech bounded by consecutive pauses) was also found to be a domain of phrase-final lengthening. Note that a finding such as Dankovičová's, i.e. a progressive reduction of the articulation rate in the Swedish prosodic phrase, cannot be interpreted as evidence in favor of the F0 dependence explanation. In the revised version of the Lund intonation model (Bruce 1982a, 1982b and 1984), the downward trend of F0 within the prosodic phrase is described with downstep. Downstep (as modeled in the revised Lund model) ensures that the accents' F0 excursions become increasingly smaller as the phrase progresses. A dependency of segment durations on F0 excursion size would therefore predict a general/progressive increase of the articulation rate rather than a decrease.

Dankovičová's (1997) method will not allow us to pinpoint the exact domain of final lengthening, if indeed final lengthening exists in Southern Swedish. Nevertheless, it will allow us to answer the question of whether or not there is a lower articulation rate in phrase-final words, and if so, whether the reduction in articulation rate begins before the last prosodic word of the phrase or not.

### 3.1.5 Research question

The research question to be answered in the present chapter concerns the existence of articulation rate variation within the prosodic phrase in Southern Swedish. Is there variation in articulation rate within the phrase, and if so, what is the nature of this variation? Two hypotheses will be tested: 1) that the position of a word has an effect on the articulation rate, and 2) that the articulation pattern demonstrates a progressive slowing down.

## 3.2 Method

### 3.2.1 Speech material

The speech material used in the present study has been extracted from the SweDia *2000* database (Bruce, Elert, Engstrand, Eriksson and Wretling 1999). Within *SweDia 2000*[5] (*Phonetics and Phonology of the Swedish Dialects around the Year 2000*) speech samples from over 100 Swedish dialects were collected during 1998-2000. In the database, each dialect is represented by at least 12 speakers (three young women (between the ages 20 and 30), three elderly women (between the ages 55 and 75), three young men and three elderly men). The recorded material consists of both spontaneous speech and words and phrases elicited with a number of specific research questions in mind (e.g. the prosodic typology of the Swedish dialects, see Bruce *et al.* 1999). The speech material was recorded by project assistants who got in contact with the subjects mainly through centers that study local geography, history and folklore (*hembygdsföreningar*). The recordings were made in the homes of the subjects in order to obtain natural dialect recordings. Care was taken to turn off disturbing sound sources in the recording environment. The technique used by the project assistants when approaching the subjects is described in some detail in Aasa *et al.* (2000). The subjects were informed orally of the aims of the *SweDia 2000* project, and they were also guaranteed full anonymity. They were not paid for their participation but were given a small gift. The recordings were made with a sampling rate of 44.1 kHz and 16 bit resolution, and transferred digitally to a Sun workstation and stored as ESPS/Waves+™ files. The sampling frequency has subsequently been converted to 16 kHz. In the present chapter, speech extracted from the spontaneous part of the database is analyzed.

---

[5] *SweDia 2000* is supported by the Bank of Sweden Tercentenary Foundation, The Cultural Foundation (Grant No. 1997-5060:01-02).

The present study on articulation rate variation was undertaken in two steps. First, a subpart of the data, the speech of five speakers, was analyzed in order to get an idea of whether or not the method used to measure phrase-final lengthening (i.e. as a change in articulation rate in the final part of the prosodic phrase) was suitable for investigating final lengthening in Southern Swedish. Secondly, after having determined that final lengthening could be observed in our data with the method chosen, the data was expanded to include speech from ten speakers. As a consequence of the results of the first analysis, however, the data was first re-segmented.

In the first step of the study, the speech of three female speakers (one from the younger generation of speakers recorded, and two from the older generation) and two male speakers (from the older generation) was analyzed (speakers 1, 2, 4, 5 and 6[6]). Subsequently, in the second step of the study, the speech of two additional female speakers (from the younger generation) and three male speakers (from the younger generation) was included (speakers 3, 7, 8, 9 and 10), see Table 3.2. Care was taken to include speech from speakers from all five recording locations in *Skåne*, and both male and female subjects from the younger and older generation.

**Table 3.2** *Presentation of the speakers (1-10)*

| | *Bara* | *Broby* | *Löderup* | *Norra Rörum* | *Össjö* |
|---|---|---|---|---|---|
| *Younger generation (20-30 ys)* | Male speaker (7) Female speaker (8) | Male speaker (10) | Female speaker (3) | Male speaker (9) | Female speaker (6) |
| *Older generation (55-75 ys)* | | Female speaker (1) | Male speaker (2) | Female speaker (4) | Male speaker (5) |

The first 100 prosodic phrases in the speakers' spontaneous recordings were initially chosen for the analysis. Only those phrases without phrase internal pauses or fillers were investigated. Disfluent phrases were excluded since they may contain segmental lengthening associated with another domain than the prosodic phrase (see Dankovičová 1997). Prosodic phrases containing fillers like [ɛ] (filled pauses)

[6] The results of the first step of this study have been reported elsewhere (Hansson 2002). In Hansson (2002), the speech of six speakers (speakers 1-6) was investigated. However, due to a segmentation error found in the data from speaker 3, her speech is excluded from the results presented in section 3.3 (but included in the second step of the study, see section 3.4).

were excluded since it was not clear to us if they were best counted as syllables of their own or not. It would seem that many words are lengthened either by increasing the duration of their segments or by adding an [ɛ] to the end of the word (or a combination of both strategies). The alternative way of increasing the length of a word by adding an [ɛ] was observed and analyzed as such already by Gårding in 1967b. When the articulation rate in a word is expressed in actually occurring syllables per second, then the observed articulation rate differs greatly between two equally long instances of the same word depending on the strategy used to lengthen it. A word like *och* /ɔk/ 'and', e.g., has twice the articulation rate when lengthened through [ɛ] addition, [ɔk$^{h}$ɛː] or [ɔɛː], than by prolongation of existing segments, [ɔː]. The difference in articulation rate does not reflect the auditory impression which is one of clear lengthening in both cases. However, since care was taken in all other cases to count the actual number of syllables in each prosodic word (rather than the numbers of syllables the word contains in its citation form) we excluded phrases with fillers from our study.

The observed strategies are, nevertheless, interesting to study further as they may reflect problems in speaking, like the pronunciation of *the* in English. Fox Tree and Clark (1997) have found that *the*, which is normally pronounced [ðĕ] or [ðə] before consonants and [ði] before vowels, is pronounced with a nonreduced vowel before both consonants and vowels when used to signal an immediate suspension of speech. The English fillers 'uh' and 'um' also have a distribution that can be related to problems in speaking; 'uh' signals a short interruption and 'um' a more serious one (Clark 1994). Swerts (1998) has reported on findings suggesting that Dutch fillers have a specific distribution in discourse. He found that filled pauses are more frequent at so-called major discourse boundaries in spontaneous monologues than in other positions.

### 3.2.2 Segmentation criteria and measurements

The positions of prosodic phrase boundaries and word boundaries as well as the number of syllables contained in each word were marked in label tiers by the author after listening and visually inspecting F0 traces and broadband spectrograms of the speech material in the speech analysis program ESPS/Waves+™.

For the analysis of the articulation rate variation within the prosodic phrase, the articulation rate (measured in syllables per second) was calculated for each prosodic word (the defining feature of the prosodic word in Swedish being that it contains an accent, see section 1.4). Swedish is not a language with a fixed stress position,

but the by far most common stress pattern in our data (counting tokens) involves an initial stress. Therefore, the stressed syllables were chosen to serve as landmarks in the segmentation of the phrases. Only in phrase-initial position were unstressed syllables attached to the following stressed syllable instead of the preceding. In order to decide the exact positions of the prosodic word boundaries (the onsets of the stressed syllables), the 'Maximal Onset Principle' (which puts consonants preferentially into onsets rather than codas) was applied. In counting the number of syllables in each prosodic word, care was taken to count the actual number of syllables, rather than the number of syllables the word contains in its citation form. Figure 3.2 shows an example of how the prosodic phrases were segmented and the label tiers used.

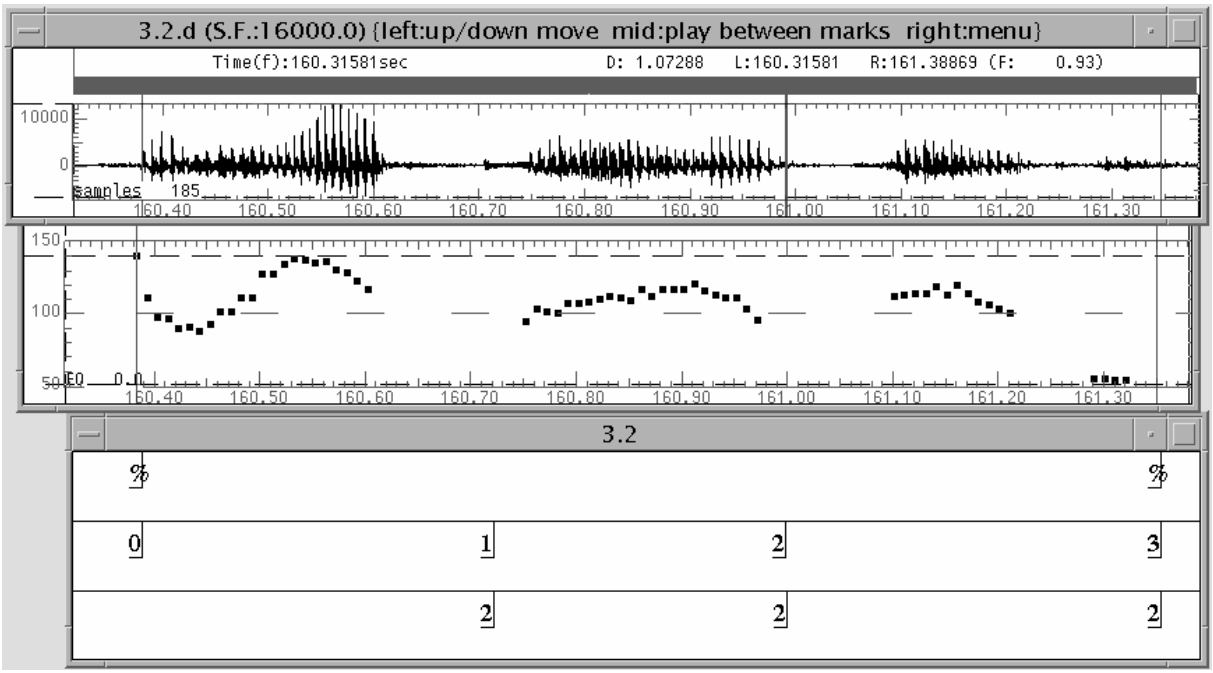


**Figure 3.2** *Waveform and F0-contour of the prosodic phrase* och ˈhaft ˈhèla ˈtiden *'and have had all along' (male speaker). The three label tiers are, from top to bottom: 1) a phrase tier with the beginning and the end of the prosodic phrase marked with '%', 2) a word tier with labels for the beginning (0) of the initial word as well as the end of the initial (1) and possible following words, and 3) a syllable tier with labels for the number of syllables within the words in the word tier.*

The information in the labeling tiers on the length of each word, its position in the prosodic phrase and size (as expressed by the number of syllables it contains) was imported to the computer statistics package SPSS[7]. There, the one-word phrases were excluded (since no analyses of articulation rate changes within one-word phrases can be done with the chosen method). Then, 2-, 3-, 4-, 5- and 6-word

[7] SPSS 10.0 for Macintosh.

phrases were grouped separately, and the articulation rate in each word was calculated by dividing the number of syllables the word contained with its length (in seconds). The subsequent statistical analyses were also undertaken using SPSS.

In the second step of the study, the data was re-segmented as it was found that the method described above was not suitable for measuring the articulation rate in phrase-initial words as labeled first. The re-segmentation is described in detail in section 3.3.1.

## 3.3 Results from the analysis of a subpart of the data

There are 278 prosodic phrases containing more than one prosodic word in the subpart of the material investigated first. The distribution of 2-, 3-, 4-, 5- and 6-word phrases for each speaker (1, 2, 4, 5 and 6) separately and all speakers pooled together is shown in Table 3.3. Due to the small number of 5- and 6-word phrases in the material, the following exposition of the results is based mainly on the examination of phrases containing 2 to 4 words.

**Table 3.3** *Number of prosodic phrases containing two to six prosodic words produced by speakers 1, 2, 4, 5 and 6*

| Speaker | 2-word phrases | 3-word phrases | 4-word phrases | 5-word phrases | 6-word phrases |
|---|---|---|---|---|---|
| 1 | 20 | 21 | 8 | 5 | 1 |
| 2 | 17 | 21 | 8 | 2 | 1 |
| 4 | 18 | 20 | 9 | 5 | 0 |
| 5 | 39 | 20 | 7 | 1 | 0 |
| 6 | 24 | 23 | 8 | 0 | 1 |
| Total: | 118 | 105 | 40 | 13 | 3 |

The idea proposed in Dankovičová (1997), is that the articulation rate variation within the prosodic phrase follows a certain pattern, a slowing down or reduction of the articulation rate. In order to test whether this is also the case in Southern Swedish, we have chosen to first rank-order the words in the phrases according to their articulation rate.

As shown in Table 3.4, 105 (89%) of the 118 2-word phrases show an AB pattern, i.e. a reduction of the articulation rate where the first word (A) is articulated with a higher articulation rate than the second and final word (B). All speakers use the AB pattern more frequently than the BA pattern.

**Table 3.4** *Ordinal patterns in 2-word phrases*

| Speaker | AB | BA |
|---|---|---|
| 1 | 19 | 1 |
| 2 | 14 | 3 |
| 4 | 16 | 2 |
| 5 | 32 | 7 |
| 6 | 24 | 0 |
| Total: | 105 | 13 |

As shown in Table 3.5, 54 (51%) of the 105 3-word phrases demonstrate an ABC pattern, i.e. a progressive slowing down of the articulation rate where the first word (A) is pronounced faster than the second word (B), which in its turn is pronounced faster than the third and final word (C). Another 31 phrases (30%) show a reduction of the articulation rate that is observable only in the comparison of the articulation rates in the second and third word, i.e. a BAC or CAB pattern. Evidence of final lengthening is, in other words, found in 85 of the 105 phrases (i.e. 81%). No phrase demonstrates a CBA pattern, i.e. a pattern in which the articulation rate gradually increases.

**Table 3.5** *Ordinal patterns in 3-word phrases*

| Speaker | ABC | ACB | BAC | BCA | CAB | CBA |
|---|---|---|---|---|---|---|
| 1 | 10 | 5 | 6 | 0 | 0 | 0 |
| 2 | 10 | 6 | 5 | 0 | 0 | 0 |
| 4 | 13 | 2 | 4 | 0 | 1 | 0 |
| 5 | 10 | 2 | 7 | 0 | 1 | 0 |
| 6 | 11 | 2 | 7 | 1 | 2 | 0 |
| Total: | 54 | 17 | 29 | 1 | 4 | 0 |

As shown in Table 3.6, the most common pattern in the 4-word phrases is the ABCD pattern, i.e. a gradual decrease in articulation rate over the course of the phrase. The ACBD pattern is also relatively common. In the ACBD pattern, the reduction of the articulation rate is visible in the last two words. It is also clear that in the majority of the 4-word phrases, it is either the first or the second word in the phrase that is pronounced the fastest (in 33 cases of 40), and that the phrase-final word is either the word with the lowest or second lowest articulation rate in the phrase (in 33 cases of 40). Evidence of phrase-final lengthening, i.e. a reduction of the articulation rate between the penultimate and final word is found in 29 of 33 phrases (i.e. 88%).

**Table 3.6** *Ordinal patterns in 4-word phrases (only patterns existing in the data included)*

| Pattern | Number of phrases | Pattern | Number of phrases |
|---|---|---|---|
| ABCD | 11 | BADC | 2 |
| ABDC | 2 | BCAD | 3 |
| ACBD | 6 | BCDA | 1 |
| ACDB | 1 | BDCA | 1 |
| ADBC | 2 | CABD | 2 |
| ADCB | 1 | CADB | 3 |
| BACD | 3 | CBAD | 2 |

The reduction of the articulation rate over the course of the phrase is also observable in the mean articulation rates in the material, as shown in Table 3.7. Typically, the articulation rate in the phrase-initial words is about 7 syllables per second and in the final word about 4 syllables per second. The largest differences in mean articulation rate between successive words are found in the comparisons between the articulation rate in the phrase-final and penultimate words.

**Table 3.7** *Mean articulation rate (in syllables per second) and standard deviations (in syllables per second) for the words in 2-, 3-, 4- and 5-word phrases*

| | 2-word phrases (n=118) | | 3-word phrases (n=105) | | 4-word phrases (n=40) | | 5-word phrases (n=13) | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 1st word | 6.9 | 2.1 | 6.6 | 1.5 | 7.3 | 2.4 | 7.1 | 1.6 |
| 2nd word | 4.1 | 1.4 | 5.8 | 1.6 | 6.0 | 1.6 | 5.9 | 1.3 |
| 3rd word | | | 4.2 | 1.3 | 5.5 | 1.8 | 5.9 | 2.0 |
| 4th word | | | | | 3.7 | 1.2 | 5.6 | 1.5 |
| 5th word | | | | | | | 4.3 | 1.5 |

In order to test the hypothesis that a word's position in the phrase has an effect on its articulation rate, a GLM (general linear modeling) procedure was used. The dependent variable was articulation rate, and the two factors were position (word's position in phrase) and speaker. Each phrase type (the 2-, 3- and 4-word phrase) was analyzed separately.

The factor of position was significant for all phrases: 2-word phrases ($F(1, 226)=154.4$, $p<.001$), 3-word phrases ($F(2, 300)=74.6$, $p<.001$) and 4-word phrases ($F(3, 140)=30.1$, $p<.001$). No analyses were made of the 5- and 6-word phrases because of the small number of occurrences. In order to analyze the pattern of

articulation rate variation, i.e. if a pattern of progressive slowing down exists, a posthoc Tukey test was done on the means. In the 3-word phrases the mean articulation rate of all three words was significantly different from each other at the .01 level. In the 4-word phrases, on the other hand, significant differences in mean articulation rate were found only in the comparisons of the mean articulation rates in words 1 and 2, 1 and 3, 1 and 4, 2 and 4, and 3 and 4 ($p<.01$). In other words, the only significant differences in articulation rate between two successive words were found in the comparison between first and second word in the phrase and between the penultimate and final word.

The factor of speaker was not significant at the .01 level in any of the phrase types, nor was the two-way interaction position by speaker, indicating that the speakers did not differ in their patterns of articulation rate. This latter finding is consistent with the results of the rank ordering of the words according to their articulation rate.

### 3.3.1 Discussion

The first analysis of speech from five speakers revealed a significant effect of word position on articulation rate. Like Dankovičová (1997), we also found some evidence to suggest a progressive reduction of the articulation rate over the phrase, i.e. in the rank ordering of the words according to their articulation rate and the mean articulation rates in words in different positions in the phrase. However, regardless of the length of the phrase (as expressed in the number of words it contains), the articulation rate in the phrase-initial word was roughly the same in phrases of all lengths as was the articulation rate in phrase-final words. The difference in mean articulation rate between words in the phrase was small in the 4- and 5-word phrases, and the reduction in articulation rate between all successive words in the phrase was only significant in the 2- and 3-word phrases. This fact may indicate that the reduction in articulation rate is not progressive but rather that it is the phrase-initial and phrase-final words that differ in articulation rate from the remaining words of the prosodic phrase. The finding that the phrase-final words were more slowly articulated than preceding words in the phrase is not surprising. That the largest difference in mean articulation rate was found in the comparison of the articulation rate in phrase-final and penultimate words indicates that final lengthening exists in Southern Swedish. However, the finding that phrase-initial words were more quickly articulated than the following words is surprising, and may indicate that we need to consider the factor of word size (as expressed in number of syllables the word contains) in our analysis of the articulation rate

variation. Although it was shown in Dankovičová (1997) that size is not a reliable determinant of a word's articulation rate, the fact that the phrase-initial words in our data tend to contain more syllables than non-initial words (as a consequence of the segmentation criteria we used, see section 3.2.2), makes it is important to pay some extra attention to the phrase-initial words' articulation rates. Stressed syllables are considerably longer than unstressed syllables, and consequently the larger the number of unstressed syllables in the word, the faster the measured articulation rate (in syllables per second).

When adding another five speakers to our analysis, we therefore chose to re-segment our speech material in such a way that the unstressed syllables preceding the first stressed syllable are excluded from the calculation of the articulation rate in phrase-initial words. In the example given in Figure 3.2 above, the label 0 in the middle tier (which indicates the start of the phrase-initial word) was moved from the beginning of the phrase to the onset of the first stressed syllable in the re-segmentation of the data. The number of syllables recorded in the bottom tier was consequently also changed (from two to one in this particular example).

As we add more material to our investigation, we also make analyses of 5-word phrases possible.

## 3.4 Results after re-segmentation and addition of more data

There are 518 prosodic phrases without disfluencies, fillers and phrase-internal pauses that contain more than one prosodic word in the speech material as a whole.

As shown in Table 3.8, 182 (82%) of the 221 2-word phrases show an AB pattern, i.e. a reduction of the articulation rate where the first word (A) is articulated with a higher articulation rate than the second and final word (B). All ten speakers use the AB pattern much more frequently than the BA pattern.

**Table 3.8** *Ordinal patterns in 2-word phrases*

| Speaker | AB | BA |
|---|---|---|
| 1 | 18 | 2 |
| 2 | 14 | 3 |
| 3 | 12 | 4 |
| 4 | 13 | 5 |
| 5 | 32 | 7 |
| 6 | 23 | 1 |
| 7 | 18 | 5 |
| 8 | 18 | 2 |
| 9 | 22 | 9 |
| 10 | 12 | 1 |
| Total: | 182 | 39 |

As shown in Table 3.9, 74 (41%) of the 180 3-word phrases demonstrate an ABC pattern, i.e. a progressive slowing down of the articulation rate. Another 54 phrases (30%) show a reduction of the articulation rate that is observable only in the comparison of the articulation rates in the second and third word, i.e. a BAC or CAB pattern. In other words, evidence of final lengthening is found in 71% of the phrases.

**Table 3.9** *Ordinal patterns in 3-word phrases*

| Speaker | ABC | ACB | BAC | BCA | CAB | CBA |
|---|---|---|---|---|---|---|
| 1 | 11 | 4 | 4 | 1 | 1 | 0 |
| 2 | 10 | 5 | 4 | 1 | 0 | 1 |
| 3 | 5 | 3 | 2 | 1 | 1 | 1 |
| 4 | 11 | 1 | 6 | 0 | 0 | 2 |
| 5 | 8 | 1 | 6 | 3 | 2 | 0 |
| 6 | 9 | 2 | 8 | 2 | 2 | 0 |
| 7 | 5 | 5 | 2 | 2 | 1 | 1 |
| 8 | 3 | 2 | 5 | 2 | 1 | 0 |
| 9 | 4 | 2 | 3 | 1 | 0 | 1 |
| 10 | 8 | 4 | 6 | 2 | 0 | 2 |
| Total: | 74 | 29 | 46 | 15 | 8 | 8 |

In the 4-word phrases, a progressive reduction of the articulation rate, i.e. an ABCD pattern, can be observed in 15 of the 78 phrases. Another 15 phrases have a progressive reduction of the articulation rate observable only in the final three words. A total of 58 (74%) phrases show evidence of phrase-final lengthening (a

reduction of the articulation rate observable in the comparison of the penultimate and final word). In the 5-word phrases, a reduction of the articulation rate between the penultimate word and the final word is found in 26 of the 33 phrases (i.e. 79%). No examples of progressive articulation rate reduction (an ABCDE pattern) exist in the 5-word phrases.

The articulation rate changes in the different phrase types can also be observed in the mean articulation rates, as shown in Table 3.10. The articulation rate in the phrase-initial words is about 6.5 to 7 syllables per second and in the final word about 4 to 4.5 syllables per second. The phrase internal words have articulation rates of about 5.5 to 6 syllables per second. Thus, despite the re-segmentation of the data, some evidence of articulation rate changes occurring in other parts than the final part of the phrase has been found. The largest differences in mean articulation rate between successive words are nevertheless found in the comparisons between the articulation rate in the phrase-final and penultimate words.

**Table 3.10** *Mean articulation rate (in syllables per second) and standard deviations (in syllables per second) for the words in 2-, 3-, 4- and 5-word phrases*

| | 2-word phrases (n=221) | | 3-word phrases (n=180) | | 4-word phrases (n=78) | | 5-word phrases (n=33) | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 1st word | 6.7 | 2.2 | 6.5 | 2.0 | 6.8 | 2.9 | 7.1 | 2.0 |
| 2nd word | 4.3 | 1.6 | 5.8 | 1.6 | 6.1 | 1.6 | 6.0 | 1.7 |
| 3rd word | | | 4.5 | 1.5 | 5.6 | 1.8 | 6.4 | 1.9 |
| 4th word | | | | | 4.1 | 1.4 | 6.0 | 1.4 |
| 5th word | | | | | | | 4.5 | 1.4 |

The hypothesis that a word's position in the phrase has an effect on its articulation rate was also tested on the material as a whole using a GLM procedure. The dependent variable was once again articulation rate, and the two factors were position (word's position in phrase) and speaker. Each phrase type (the 2-, 3-, 4- and 5-word phrase) was analyzed separately.

The factor of position is significant for all phrases: 2-word phrases ($F(1, 422)=172.8$, $p<.001$), 3-word phrases ($F(2, 510)=68.3$, $p<.001$), 4-word phrases ($F(3, 272)=26.2$, $p<.001$) and 5-word phrases ($F(4, 120)=7.1$, $p<.001$). A posthoc Tukey test was done on the means to examine if a pattern of progressive slowing down exists. In the 3-word phrases the mean articulation rate of all three words is significantly different from each other at the .01 level. In the 4-word phrases, on

the other hand, significant differences in mean articulation rate have been found only in the comparisons of the mean articulation rates in words 1 and 3, 1 and 4, 2 and 4, and 3 and 4 ($p<.01$). In other words, the only significant difference in articulation rate between two successive words can be observed in the comparison between the final and penultimate word. In the 5-word phrases, significant differences in mean articulation rate were found only in the comparisons of the mean articulation rates in words 1 and 5, 2 and 5, 3 and 5, and 4 and 5 ($p<.01$), i.e. between successive words only in final position.

The factor of speaker is not significant at the .01 level in any of the phrase types. Some differences significant at the .05 level were found, but the finding that different speakers use different articulation rates is not unexpected. The two-way interaction position by speaker is not significant at the .01 level in any of the phrases. This indicates that the speakers, despite some general differences in articulation rates, do not differ in their patterns of articulation rate.

## 3.5 Summary and discussion

In the first step of this study on articulation rate variation, speech from five speakers was investigated and the results revealed a significant effect of word position on articulation rate in 2-, 3- and 4-word phrases. Some evidence to suggest a progressive reduction of the articulation rate over the phrase was also found. However, it appeared that the observed progressive reduction was a consequence of the fact that the phrase-initial and phrase-final words differed in articulation rate from phrase-internal words. In other words, the evidence of progressive reduction in phrase-internal words was weak.

The finding that the largest difference in mean articulation rate was found in the comparison of the articulation rate in phrase-final and penultimate words indicates that final lengthening exists in Southern Swedish. However, the finding that phrase-initial words were more quickly articulated than the following words was seen as a possible consequence of the segmentation criteria used. The speech material was consequently re-segmented, and the speech of another five speakers was included. By including more data, we made it possible to also study 5-word phrases that, due to too few occurrences in the subpart of the material, were not included in the first step of the study.

In the second step of the study, a significant effect of word position on articulation rate was found in the 2-, 3-, 4- and 5-word phrases. Despite the fact that the data

had been re-segmented, phrase-initial words were still more quickly articulated than phrase-internal and phrase-final words. In the 3-word phrases, a progressive articulation rate reduction could therefore still be observed. Nevertheless, the analysis of the 4- and 5-word phrases showed that significant differences in articulation rate between successive words are only found in the final part of the prosodic phrase (between the final and penultimate word).

The perceptual relevance of the fast articulation rate in phrase-initial words is worth discussing. The results of the present study can be interpreted as indicating that the perceptually relevant aspect of duration cues in prosodic phrasing is not only the lengthening of segments phrase-finally, but also the apparent shortening of segments phrase-initially. It is reasonable to assume that the difference in articulation rate between the phrase final word and the phrase initial – being larger than the difference in articulation rate between the penultimate and phrase-final word – is a strong marker of the presence of a phrase boundary. In particular, we may speculate that this is the case for prosodic phrases not marked by pauses. However, more investigations of phrase-initial shortening are needed. The accents occurring in phrase-initial words are often less prominent than accents occurring in the final part of the phrase (i.e. non-focal)[8]. The explanation for this can be found in the information structure of the phrases. The 'new' or focal information is placed in the final part of the phrase whereas the 'old' information already known to the listener (the 'theme' or 'topic') is placed at the beginning of the phrase (see section 1.4). New information tends to be prosodically prominent (i.e. associated with focal accents) whereas old information is more frequently associated with non-focal word accents. One of the features of focal accents is the segmental lengthening they give rise to (Bruce 1981, Heldner 2001)[9]. It is possible that listeners do not perceive phrase-initial words as more rapidly articulated than phrase-internal words as they do not expect any lengthening resulting from focal accentuation at the beginning of phrases. In order to answer the question of whether or not such a phenomenon as phrase-initial shortening exists and is communicatively relevant in Swedish, our study needs to be complemented by perception studies.

[8] Note that we by 'phrase' mean 'prosodic phrase' here. Although the focal accent in Southern Swedish tends to be placed earlier in e.g. noun phrases than in Stockholm Swedish (see section 3.1.3), it rarely occurs in prosodic phrase-initial position.

[9] On the other hand, experiments with Dutch stimuli have shown that less complex intonation contours are perceived as faster articulated than more complex contours (Rietveld and Gussenhoven 1987).

Based on the results at hand, we can nevertheless draw the following conclusions: Firstly, there is no doubt that phrase-final lengthening exists in Southern Swedish despite the dialect's prosodic similarities with Danish, and the lack of a (high) phrase accent. We cannot pinpoint the exact domain of the phrase-final lengthening, but it would appear to affect only the final word of the phrase. Final words are articulated slower then penultimate words. Penultimate words, on the other hand, are not more slowly articulated than preceding (non-initial) words. Secondly, the rank ordering of our data reveals that final lengthening is used in about 80% of the phrases. We interpret this as suggesting that duration cues are just as important cues to prosodic phrase structure in spontaneous speech as they are in read speech (see Bruce *et al.* 1993).

CHAPTER 4

# Tonal coherence within the prosodic phrase

## 4.1 Introduction

In the present chapter, we shift our focus of attention from the durational signaling of phrase boundaries to the tonal means used in speech to signal coherence within the prosodic phrase.

The Lund model for intonation exists in two versions. The main difference between the two is related to the modeling of the downward trend of F0 within the prosodic phrase. In the version advocated by Eva Gårding (see e.g. Bruce and Gårding 1978, Gårding 1983, Gårding and Bruce 1981), time-dependent declination is assumed, whereas in the revised Lund intonation model (Bruce 1982a, 1982b, and 1984), downstep is assumed[1]. Our main objective in this

[1] Ladd (1983) has proposed the terms 'Contour Interaction' and 'Tone Sequence' to refer to the two general, existing approaches to the description of the downward trend of F0. In the so-called 'Contour Interaction' approach (Ladd 1983: 40), "the pitch contour of an utterance is specified by a number of separate components which generate, for prosodic domains of various sizes, pitch configurations that are superimposed or overlaid on one another". In other words, the F0 contour is thought to consist of a global downslope on which the accents are superimposed. In the 'Tone

chapter is to investigate whether the downward trend of F0[2] in a spontaneous speech material is best described within the original Lund model (by assuming time-dependent declination) or the revised Lund model (by assuming downstep). Although we follow Bruce in calling the model that he advocates the revised version of the Lund model, it is important to stress that the original version of the Lund model was not abandoned with the introduction of the revised model (see e.g. Touati 1988), and that, just as in discussions of downstep in English, "the debate between the two approaches is far from being settled" (Nolan 1995: 253). Furthermore, the two approaches have different implications for the amount of preplanning needed, which makes a comparison of the two particularly interesting in a study of spontaneous speech.

### 4.1.1 Coherence within the prosodic phrase

There are two reasons for our interest in studying the downward trend of F0. The first is the possible insights that can be reached as to the amount of preplanning involved in prosodic phrasing, and the second is the belief that the downslope of F0 very well could be one of the most important means used to signal coherence within the prosodic phrase. Regardless of what approach is taken to the description of the downward trend of F0, the downsloping nature of F0 is believed to signal coherence between prosodic words that are grouped together in a prosodic phrase.

In the same manner as a reset of F0 is believed to signal the presence of a prosodic phrase boundary, a gradual or stepwise lowering of F0 over the course of the prosodic phrase can be regarded as an active signal of coherence. Prosodic phrase-internal pauses are perceived as phrase-internal due not only to the lack of boundary signaling cues (e.g. lack of boundary tones and F0 resets, see Figure 4.1), but quite possibly also to the more active coherence signaling realized by accent peak lowering.

---

Sequence' approach (Ladd 1983: 40) by contrast, one assumes "no layer or component of intonation separate from accent: intonation consists of […] a sequence of tonal elements".

[2] In what follows, we will use the terms 'F0 downtrend/downward trend of F0' and 'downslope (of F0)' to refer to the empirically observable tendency for the fundamental frequency to slowly drop, regardless of whether it is assumed to be the result of time-dependent declination or downstep. We will only use the term '(time-dependent) declination' for such downslope of F0 which is thought to be the result of a global time-dependent downslope.

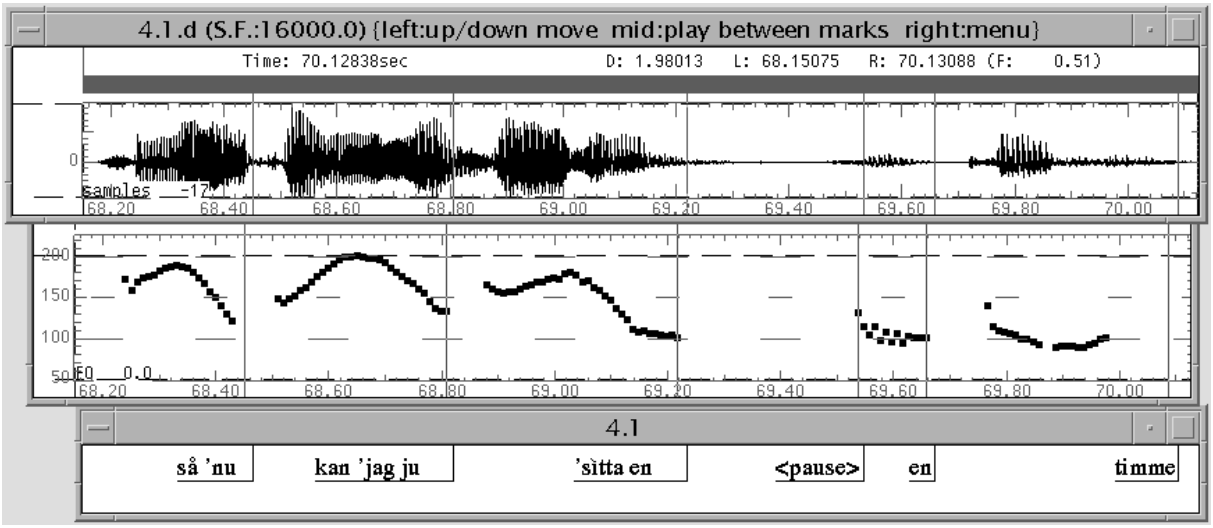


**Figure 4.1** *Speech wave and F0 contour of the prosodic phrase* så nu kan jag ju sitta en (p) en timme *'so now I can sit down for an (p) an hour' (Oss_om[3]). Since there is no accent to be subjected to downstep after the phrase-internal pause, the F0 value of the last accent L before the pause is maintained.*

Thus, when listening to prosodic phrases containing phrase-internal pauses, the change in F0 across the pause (if any) appears to have great importance for whether or not we perceive the speech following the pause as a continuation of what was previously said or as the beginning of something new[4].

Our expectation of a lowering of the accent peaks in the phrase is strong enough for us to perceive the second accent in Figure 4.2 as more prominent than the first accent although their peaks are of equal height.

That listeners normalize for the downward trend of F0 when judging the relative prominence of accents, has been demonstrated by Pierrehumbert (1979) in a series of experiments with English stimuli, and this would appear to be equally true for Southern Swedish listeners. In Southern Swedish, prominence is signaled with pitch range (rather than with an extra tonal gesture as in Stockholm Swedish, see section 1.4).

[3] *Oss_om* stands for speech from a **m**ale speaker from **Öss**jö (the **o**lder generation), see section 4.2.1.

[4] Other features, such as the presence or absence of prepausal lengthening, may naturally also be important.

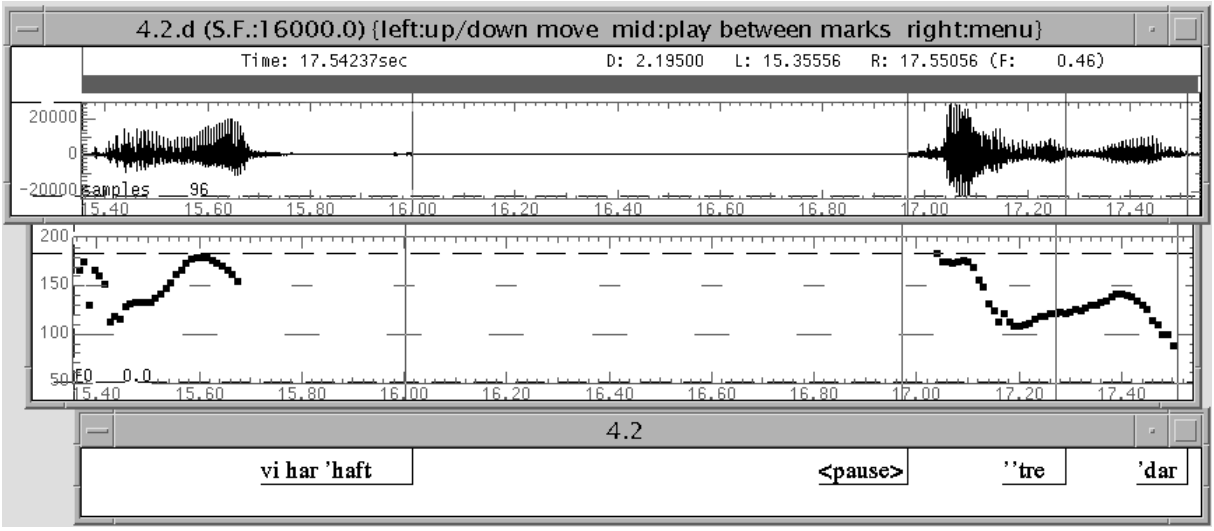


**Figure 4.2** *Speech wave and F0 contour of the prosodic phrase* vi har haft (p) tre dar *'we have had (p) three days' (Oss_om).*

After the following introduction to how the downward trend of F0 is modeled assuming either time-dependent declination or downstep, we will address the question of how speakers go about producing this signal of coherence in spontaneous speech.

## 4.1.2 Time-dependent declination

The term 'declination' was coined by Cohen and 't Hart in the 60's to refer to a phenomenon that was commented on by Pike already in 1945, namely, the tendency for F0 to gradually decline over the course of an utterance (Collier, Cohen and 't Hart 1982). This tendency of the fundamental frequency to slowly drop over the utterance has been observed in numerous languages and has even been claimed to be a language universal (Bolinger 1962). With the claim of declination being a language universal in mind, it is justified to look for a physiological explanation to its existence. One such explanation has been discussed by among others Collier (1975), who suggests that a decrease in subglottal air pressure during the utterance is responsible for the declination phenomenon. However, other studies have indicated that the decrease in subglottal air pressure is not large enough to account for all the F0 drop (Maeda 1976) and objections against the universality of declination have also been put forward on the grounds that declination is not always present in speech, particularly not in spontaneous speech (Lieberman, Katz, Jongman, Zimmerman and Miller 1985, Lieberman 1986). A third perspective is taken in Gussenhoven (2002) where the universality of intonational meaning is claimed to be based on so-called 'biological codes', metaphors of biological conditions that influence the speech production process.

One of these codes, the 'Production Code', associates high pitch with the utterance beginning and low pitch with its end due to a correlation between 'breath groups' (Lieberman 1967) and utterances, and between drop in subglottal air pressure and drop in F0. The downward trend of F0 is then grammaticalised in different ways in different languages.

The introduction of the concept declination made it possible for Cohen and 't Hart to interpret intonation contours as consisting of two components: a declination line and a set of F0 movements for which the declination line serves as a reference (Collier *et al.* 1982). This way of representing the intonation contour has since then been used in the description of a number of different languages' intonation, e.g. English (Cooper and Sorensen 1981, Maeda 1976, O'Shaughnessy and Allen 1983, Pierrehumbert 1980), Danish (Thorsen 1980 and 1983) and Swedish (Bruce and Gårding 1978, Gårding 1983, Öhman 1967).

The original Lund model for intonation (see e.g. Bruce and Gårding 1978, Gårding 1983, Gårding and Bruce 1981), is an example of a superpositional or hierarchical intonation model. A basic assumption that underlies the model is that intonation proper can be separated from accentuation. The gradual decline in F0 is assumed to be directly accessible to empirical observation. In model terms, the word and phrase accents (represented as Hs and Ls) are inserted on reference lines, on a baseline – topline structure (the sentence intonation), and thereafter connected. The rate of declination is dependent on phrase length, as the F0 start point and end point are assumed to be constant. Thus, as often is the case in models incorporating time-dependency, longer utterances are characterized by a less steep slope than shorter ones, i.e. by length-dependency.

Note that a model incorporating time-dependent declination is not necessarily characterized by length-dependency (and constant F0 start and end points). Although claiming that the extent of declination (total F0 drop) is "more nearly constant" than the declination rate (slope of F0), Cooper and Sorensen (1981: 38) have shown that the F0 starting point varies with the length of the upcoming utterance, possibly to ensure some minimum amount of F0 slope. The sensitivity to the length of the upcoming utterance is then not reflected in the slope of F0 but in the F0 starting point. Declination is still time-dependent in the sense that F0 is thought to decline over time, and not, as in downstep models, only in connection with accents.

### 4.1.3 Downstep

Like declination, downstep, a stepwise lowering of F0, has been claimed to be present in most languages of the world. Beckman (1993: 259) notes that the work on prosody during the 80s and 90s has let us "say with a fair degree of confidence which aspects of fundamental frequency patterns are likely to generalize across languages" and one of those aspects, she claims, is that "coherence among words or phrases can be signaled when each following F0 peak is systematically reduced relative to preceding peaks", i.e. with downstep.

In the revised Lund model proposed by Bruce (1982a, 1982b, and 1984), downstep is assumed. The downstep rule proposed resembles the rule for English that Liberman and Pierrehumbert (1984) propose. In Liberman and Pierrehumbert's study, almost no time-dependent declination is found, and therefore the concept of a declining baseline that had been assumed in previous work by Pierrehumbert (1979 and 1980) is dropped. In the earlier work on downstep, the declining F0 baseline was assumed to be a characteristic of a speaker's voice. It represented the lowest F0 value the speaker was disposed to reach at any given point in the utterance[5]. F0 peaks, *p*, were scaled as the peak-to-baseline difference divided by the baseline value at the location of the peak (taking prominence relations into account), that is:

(4a) $p = (P - b_p) / b_p$

where *P* is peak height (in Hz) and $b_p$ is the baseline value (in Hz) in position *P* (Pierrehumbert 1980: 68).

In downstepped sequences of accents, the F0 value of an accent peak (H*, or the H in a bitonal accent) was furthermore thought to be subject to readjustment by downstep, that is:

(4b) $H_{i+1} = k\, H_i$

where *k* is the downstep ratio, a constant less than 1, and $H_i$ is the phonetic value of *H* (expressed in baseline units above the baseline) in position *i* (Pierrehumbert 1980: 91).

[5] Pierrehumbert's approach is an example of what Ladd (1983: 436) terms the implicit decline approach, since, in Pierrehumbert's view, it is not necessarily the F0 itself that declines, but "an abstract backdrop against which F0 is interpreted linguistically".

However, Liberman and Pierrehumbert (1984) show that the major factors shaping the F0 contour are local ones, and, consequently, that it is not necessary to assume any time-dependent declination. The downward trend of F0 is explained by a combination of a final lowering effect and the usage of stepping accents. Since the sequences of stepping accents do not converge to zero, a nonzero asymptote, *r* (the reference line), is needed in the model. The reference line, *r*, is abstract and lies somewhere between the F0 end point and the lowest peak. The following rule shows the exponential decay to the nonzero asymptote proposed by Liberman and Pierrehumbert:

(4c) $x_{i+1} - r = s\,(x_i - r)$

where *s* is a constant less than 1, and $x_i$ is peak height (in Hz) in position *i* (Liberman and Pierrehumbert 1984: 186).

Some of the main points that Bruce (1982a and 1984) incorporates in the revised Lund model are, firstly, that the overall F0 course of an utterance can be expressed in terms of the relations between successive accents. The range of the F0 accent fall decreases for successive accents (bitonal HL accents[6]) and this F0 downdrift is described by a local rule, where each F0 minimum's[7] value is a constant ratio of the preceding F0 minimum's value. The following rule describes the exponential nature of accent minima scaling in Bruce's model:

(4d) $L_{i+1} = k\,L_i$

where *k* is the downstep ratio, a constant less than 1, and $L_i$ is accent minimum value in position *i* (Bruce 1982a).

Bruce converts the F0 values of the accent minima into Pierrehumbert's (1980) baseline units above the baseline but assumes that the baseline is flat, since he has found evidence to suggest that a speaker is disposed to reach the bottom of his or her voice range already before the end of an utterance. Assuming a flat baseline means that the declining nature of F0 is entirely explained by downstep in the model[8]. Note that Bruce's assumption of a flat baseline, makes the conversion to

[6] Only H*L accents are investigated. However, it is assumed that the results can be extrapolated to the other accent, HL*.

[7] In Bruce's work, tones are identified as local accent maxima and minima.

[8] No final lowering is reported. Nevertheless, Bruce (1984: 58) recognizes that a decrease in F0 may occur on plateaus consisting of unstressed syllables but describes such a decrease as negligible.

baseline units above the baseline unnecessary in the sense that the model works equally well if the phonetic values of the accent minima are expressed in Hertz above the baseline[9]. Thus, the largest difference between Liberman and Pierrehumbert's model and Bruce's is that by referring the accent valleys' values to the reference (instead of the peaks' values), Bruce can use the baseline, the speaker's F0 floor, as reference, that is:

(4e) $L_{i+1} - b = k\,(L_i - b)$

where *k* is the downstep ratio, a constant less than 1, *b* is the baseline (in Hz), and $L_i$ is the accent minimum value (in Hz) in position *i*.

The peaks' values are not predicted very well by the rule in (4e), since they tend not to converge to the baseline, but to a higher F0 value. That this is the case, is evident from the comparison between predicted and observed values in Bruce (1982a: 107), where the predicted values of the phrase-final accent maxima are noticeably lower than the observed values. In model terms, this is solved by instead expressing the accent maxima's values as a constant interval above the preceding minimum. In other words, a reference line is indirectly built into the model, one accent rise above the baseline. It is motivated by the observation that the accent rises have the same range independently of whether they occur in the upper or lower part of a speaker's F0 range.

Bruce also addresses the question of the choice of F0 starting point. In Pierrehumbert (1980) and Liberman and Pierrehumbert (1984), the F0 starting point (the value of the first H*) is assumed to be a free choice that is governed by pragmatic or expressive factors. Bruce, however, claims a sensitivity of the F0 starting point (the F0 maximum of the first phrase accent) to the length of the utterance. He assumes that some kind of lookahead exists that "ensures that the F0 bottom of the speaker's voice will not be reached until the end of an utterance" (Bruce 1982a: 79) or rather – since downstep always ensures that values below the baseline are not generated – that a F0 value close to the baseline is reached at the end of the phrase[10]. In short phrases (i.e. in phrases with few accents) values close to

[9] Note, however, that the conversion to baseline units above the baseline has the double function of comparing values to the speaker's F0 floor and of normalizing different pitch ranges and voices of different pitches (e.g. male and female voices).

[10] Bruce (1982a) also notes that the step size of the first accent fall in an utterance is relatively constant. This implies either that the accent rises' size varies with the F0 starting point, or that the downstep constant varies with the starting point. In order for higher starting points to result in the same first step size as lower ones, then either the accent rises need to be smaller or the

the baseline are only reached if the starting point is low[11]. Thus, unlike in the model advocated by Eva Gårding, where both the F0 start and end point are assumed to be constant, no point is assumed to be truly constant in the revised Lund model. The F0 end point (the final L in the phrase) is nevertheless assumed to be the point that varies the least in the prosodic phrase.

The F0 starting point's sensitivity to sentence length is reevaluated in Bruce (1982b), where it is shown that the earlier part of a sentence's F0 contour is a "copy" of the later part of the F0 contour of the preceding sentence. In other words, the variation in F0 starting points is not dependent on any feature of the utterance, but on the utterance's placement in a larger text unit (see section 5.1.1). We will therefore not regard this (the starting point's sensitivity to sentence length) as a feature of the revised Lund model.

In Table 4.1, the main features (as regards the modeling of the downward trend of F0) of the two versions of the Lund models are summarized and compared.

**Table 4.1** *Features of the two versions of the Lund model for intonation*

| | **The original Lund model** (Bruce and Gårding 1978, Gårding 1983) | **The revised Lund model** (Bruce 1982a, 1982b, 1984) |
|---|---|---|
| Time-dependency | Yes* | No |
| Length-dependency | Yes** | Yes, a length-dependent effect is created also by downstep |
| Constant F0 starting point | Yes | No |
| Constant F0 end point | Yes | No, but the end point is assumed to be the least variable point in the F0 contour |

*A gradual decrease in F0 is observable in all parts of the utterance, including sequences of unstressed syllables. **F0 slope correlates with utterance length.

downstep constant larger. However, no attempts have been made to examine if accent rise size or the downstep constant varies with the starting point, nor is such a relationship acknowledged in the model. Instead, the model predicts that the step size of the first accent fall varies with the starting values.

[11] Bruce (1982a) reports that short utterances have higher end points than long utterances.

### 4.1.4 F0 downtrend in spontaneous speech

Much less attention has been devoted to declination and downstep in the production of spontaneous speech than in the production of read speech and in perception. To our knowledge, Swerts, Strangert and Heldner's (1996) comparison between declination in read-aloud and spontaneous speech is the only study that examines F0 downtrend in Swedish spontaneous speech.

One of the reasons for resorting to tightly controlled experiments and read speech materials when investigating declination and downstep is of course the difficulties encountered when studying spontaneous speech materials. Measurements of F0 in spontaneous speech are notoriously difficult to interpret and compare. If made in vowels, they are affected by the intrinsic F0 of the vowel in which the peak or valley is found (Lehiste 1970, Lyberg 1984, Ohala and Eukel 1987, Whalen and Levitt 1995), by the potentially F0 raising or lowering effect of a preceding and following consonant (House and Fairbanks 1953, Lehiste and Peterson 1961, Löfqvist 1975) and by the degree of prominence assigned to the accented word (Rietveld and Gussenhoven 1985). Similarily, if measured in consonants, they are affected by the F0-raising or lowering effect on F0 that particular speech sound has. These are all factors that one can control for in studies of read speech, but not in truly spontaneous speech recordings (see section 1.5). Due to the mistracking caused by vocal fry, (reliable) measurements of F0 are furthermore difficult to obtain at the end of phrases.

Liberman and Pierrehumbert (1984) have questioned the methodology used in many studies on the gradual downslope of F0. They claim that in essentially all studies made up to that point, one had attempted to measure declination on material that was not analyzed, i.e. not controlled in terms of phrasing, stress pattern and tune. Although we agree with Liberman and Pierrehumbert (1984) in much of their criticism, we nevertheless believe that it is important to try to show that the results from studies of laboratory speech apply also to spontaneous speech (if indeed they do).

In addition to the problem of how to interpret the measures of F0 valleys and peaks in uncontrolled speech, another problem when trying to measure rate of declination concerns how to fit a given F0 plot with a declination line. Results from eye-fitting procedures such as Maeda's (1976) 'visual abstraction' procedure can easily be biased by the researcher's expectations. As shown by Lieberman *et al.* (1985), different subjects (naïve subjects as well as speech researchers) examining the same sentences are not able to produce consistent results. While one person fits a given

F0 plot with a baseline that has only slight declination, another person may fit the same F0 plot with a baseline showing extreme declination. With this in mind, an alternative, quantitative method was developed and tested by Lieberman *et al.* (1985): a linear regression technique that calculates the slope of F0 using the entire F0 contour (all points).

The method developed by Lieberman *et al.* (1985) has been questioned by 't Hart (1986) who points out that the all-points regression technique, despite being claimed to be more objective than the visual abstraction procedure, does not lead to a more reliable interpretation. He gives the example of a sentence with only one pitch accent near the end. In such a sentence, the regression technique may produce a positive slope even if F0 is declining from the beginning to just before the peak ('t Hart 1986). In the reply to 't Hart, Lieberman (1986: 1840) nevertheless concludes that "it is impossible to objectively evaluate claims of declinationists since their basic measure is inherently unreplicable". He argues that they are unreplicable due to the fact that different observers of an F0 contour draw declination lines with sometimes very different slopes and notes that 't Hart offers no data to show that hand-drawn declination lines can be replicated.

In Swerts *et al.* (1996), all-points linear regression lines were used to compare declination in read-aloud and spontaneous speech. They found negative slopes in phrases and utterances in both read and spontaneous speech data. Prosodic differences between the two speaking styles were found, e.g. steeper slopes in the read than the spontaneous data, but were described as quantitative rather than qualitative. Swerts *et al.* (1996) also investigated whether or not the rate of declination was time-dependent (or rather length-dependent), i.e. if the negative slope of declination becomes more gentle as phrase length increases. Correlation coefficients between the length of phrases and utterances and the corresponding slopes were calculated and the results revealed length-dependency in both spontaneous and read speech, although lower correlations were found in the spontaneous than the read speech material. Swerts *et al.* interpreted this finding as evidence against theories in which only a minimal amount of lookahead is needed for the speaker to produce utterances adequately. However, a length-dependent effect is not sufficient to assume that declination is (time-dependent as well as) length-dependent. A length-dependent effect is created also by models including a peak-by-peak lowering (downstep).

One way to test the revised version of the Lund model's usefulness for describing spontaneous speech is by testing data to see whether slope decreases with decreasing F0 starting points. In the revised Lund model, a F0 starting point close to a speaker's baseline (F0 floor) will result in accent sequences in which each accent maximum is only marginally lower than the preceding one. As a result, the general slope of F0 will be less steep in prosodic phrases with low starting points than in phrases with high starting points, see Figure 4.3.

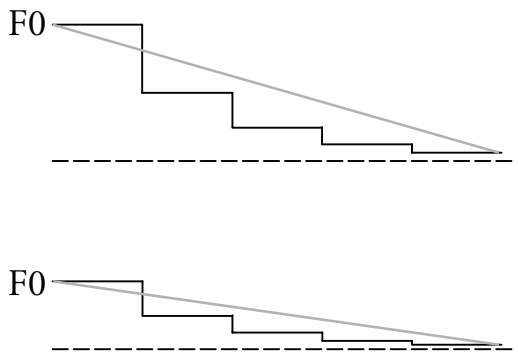


**Figure 4.3** *Schematic illustration of how step size is related to how far above the reference or base line (the broken line) the F0 starting point (the first accent's H[12]) is. The straight lines are drawn to demonstrate the resulting difference in F0 slope.*

In the time-dependent, original version of the Lund model, a relationship between F0 slope and starting point is not expected. The starting point is basically assumed to be constant and therefore no expectations exist of systematic variation in slope due to choice of starting point.

Assuming that the predictions of the time-dependent declination approach as modeled in the original Lund model and the downstep approach taken in the revised Lund model have been correctly interpreted, a correlation between F0 starting point and slope therefore provides support in favor of the downstep approach in the revised Lund model. A strong correlation between phrase length and slope would, on the other hand, provide some support in favor of the time-dependent declination approach of the original version of the Lund model, where both the F0 start point and end point are assumed to be constant, see Figure 4.4.

[12] The first phrase accent H in Stockholm Swedish.

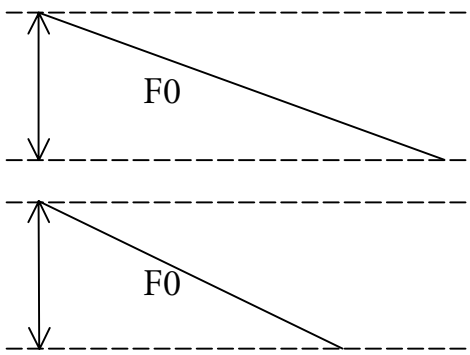


**Figure 4.4** *Schematic illustration of length-dependency, i.e. how slope and phrase length are thought to be related in models where the start and end point of the prosodic phrase's F0 contour are assumed to be constant.*

### 4.1.5 Research questions and hypotheses

We will try to answer three research questions in the present chapter.

First, we intend to determine whether evidence for length-dependent declination can be found in our data. Is the slope of F0 related to phrase length? If declination is length-dependent as described by Bruce and Gårding (1978), then we expect the negative slope to decrease with increasing phrase length. However, since a weak positive relationship between slope and phrase length are also expected to be found in models assuming downstep, a positive relationship between slope and phrase length does not unambiguously let us choose between the two models.

Secondly, we will try to determine if support for the approach taken in the revised Lund model can be found in our spontaneous data. Is the slope of F0 related to the F0 starting point? If the F0 downtrend is the result of downstep as described by Bruce (1982a), then we expect the slope to be related to the choice of F0 starting point. Large step sizes are expected to follow a high starting point (well above the speaker's baseline), and consequently a steep negative slope, whereas small step sizes and a less steep slope are expected to follow a low starting point.

The third and final research question concerns the relationship between phrase length and the speaker's choice of F0 starting point. Do speakers take the upcoming phrase's length into consideration when choosing F0 starting point? If so, then we expect phrase length to be positively correlated with starting point in our data.

## 4.2 Method

### 4.2.1 Speech material

The speech material has been extracted from the *SweDia 2000* database (see section 3.2.1 for more details concerning the database). It consists of approximately one-minute long sections extracted from the spontaneous parts of the recordings. The sections were selected by the project's assistants to be used in a public database[13]. Ten of the twenty-one one-minute sections that were initially chosen to represent *Skåne* in the database[14] have been selected for analysis. The recorded speech comes from one male subject from the younger generation (hereafter referred to as *ym*) and one male subject from the older generation (*om*) from each of the five recording locations in Skåne (Bara (hereafter referred to as *Bar*), Broby (*Bro*), Löderup (*Lod*), Norra Rörum (*Nro*) and Össjö (*Oss*)). One of the sections initially chosen for the database contained considerably more than one minute of speech (*Oss_ym*) and consequently not all speech contained in that file has been analyzed here.

### 4.2.2 Labeling of prosodic phrase boundaries

A prosodic segmentation and labeling of the material was done interactively (by the author) using the speech analysis program ESPS/Waves+™, in which the start and end of all prosodic phrases were indicated in a label tier. No distinction was made between different degrees of boundary strength, since all boundaries indicated within the Swedish transcription systems (Bruce 1994 and Bruce *et al.* 1994) (| for weakly marked boundaries, corresponding to prosodic phrases, || for strongly marked boundaries, corresponding to prosodic utterances, and ||| for extra strongly marked boundaries, corresponding to speech paragraphs) mark the end of a prosodic phrase, which is the prosodic group that interests us in the present investigation.

---

[13] See http://www.swedia.nu (Accessed 2003-01-06). In the public database, one can also listen to speech from Bjuv, a sixth recording location (test recordings).

[14] At the time when this study was carried out, the speech of two male subjects from the older generation recorded in Löderup had been chosen for inclusion in the public database. The speech that we chose to analyze was not produced by the speaker that now represents Löderup in the public database.

All phrases in the material were initially included in the study. Nevertheless, in a few cases, phrases had to be excluded due to F0 mistracking caused by vocal fry in extensive parts of the phrase.

In order to check the reliability of the labeling of the prosodic phrase boundaries in the material, two phoneticians with considerable experience in prosodic labeling (expert transcribers A and B) were also asked to transcribe the material as to prosodic grouping. They transcribed the data indicating three degrees of boundary strength as is done in the Swedish base prosody system. However, since in the present investigation we were only interested in the division of the speech into prosodic phrases, we only investigated the degree of agreement between transcribers as to phrase boundary locations. To estimate the agreement between the three transcribers (the two expert transcribers and the author), we chose to follow the approach undertaken in the ToBI evaluation (Pitrelli *et al.* 1994, Silverman *et al.* 1992) and subsequently also in the evaluation of the Swedish base prosody transcription system (Strangert and Heldner 1995b). Agreement is calculated across all possible transcriber pairs for each word in the transcription. Since there were three transcribers, there were three possible pairwise comparisons among the transcribers: the author and expert transcriber A, the author and expert transcriber B, and expert A and expert B. If all three transcribers agreed on a label for a word ('boundary' or 'no boundary'), agreement is 100%. However, if only two of the three transcribers agreed, transcriber pair word agreement is 33% (i.e. the transcribers agree in only one of the three possible transcriber comparisons). The inter-transcriber agreement reported is an average of the percentages calculated for each word in the transcription. Following the approach described above, the inter-transcriber agreement has been calculated to 93%. This is a higher index than the index reported in Strangert and Helder (1995b) for spontaneous speech (81%). However, a higher index is also expected given that we have only considered the transcribers' agreement on the locations of prosodic phrase boundaries, not their agreement on both location and strength. Based on the lack of systematic differences between the transcribers and the high inter-transcriber agreement, we feel confident in using the transcription made by the author for our investigation of F0 patterns in the prosodic phrase.

### 4.2.3 Measurements

F0 traces were generated for all files using the ESPS/Waves+™ *get f0* function. Working interactively listening to the speech and observing the F0 tracings using the speech analysis program ESPS/Waves+™, two starting points in each prosodic

phrase's F0 contour were marked in a second label tier (separate from the tier in which the prosodic phrases start and end points were marked). The F0 values of the starting points were then extracted and stored; the 'start value' (measured in the first stable part of the phrase's F0 contour) and the F0 value of the 'phrase-initial accent peak'. We have chosen to measure the F0 starting point in the first stable part of the F0 contour as a complement to the values measured in the phrase-initial accent peaks, since we cannot control for factors such as degree of prominence that affect the phonetic values of the accent peaks. No measurements were made of the phrase-initial accent valleys, however. It can be argued that it is the first accent minimum that is the relevant starting point in Bruce's revised Lund model (Bruce 1982a and b, Bruce 1984), since the low turning points' F0 values are used to calculate the values of the high turning points. However, since it is not clear whether the relevant low turning point of word accent II precedes or follows the accent's high turning point in Southern Swedish, we will only measure starting point in the phrase-initial accent peak (and the phrase's first stable part). We will return to this issue in section 4.4.1.2.

The length (in s) of each prosodic phrase was extracted automatically and stored using the labels in the label tiers (the labels marking the starts and ends of all prosodic phrases). Finally, using an all-points linear regression technique (where all the points in the F0 contour, one point each millisecond in voiced parts of the speech, were used), the slope of F0 in each prosodic phrase was calculated using the computer statistics package SPSS, see Figure 4.5. The linear regression technique was chosen rather than an eye-fitting procedure for reasons given in section 4.1.4, although the relationship between F0 and time, if it should turn out to be controlled by downstep, is not truly linear.

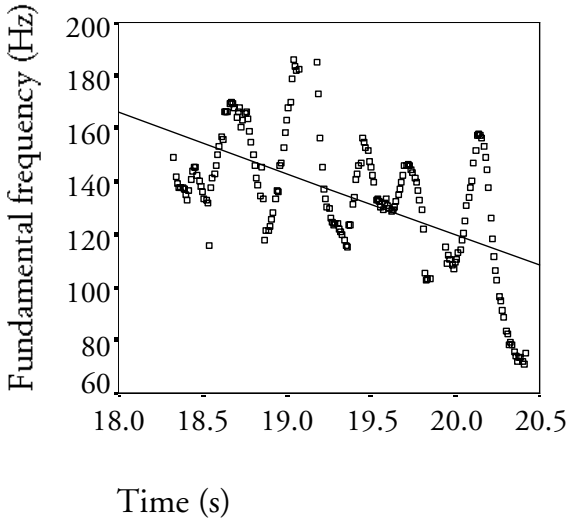


**Figure 4.5** *Scatter plot of the F0 contour of a prosodic phrase and the regression line calculated using an all-points linear regression technique (Bro_om).*

Before the F0 slope of each prosodic phrase was calculated, F0 scatter plots of all prosodic phrases were produced so that potential outliers, due mainly to F0 mistrackning, could be removed, see Figure 4.6. In the F0 scatter plot to the left, the F0 mistracking that has resulted in overly high F0 values at the end of the phrase, causes the calculated slope to reach 142 Hz/s. In the scatterplot to the right, the outliers have been removed and the calculated slope, -21 Hz/s, is more in line with the result obtained using a visual abstraction procedure.

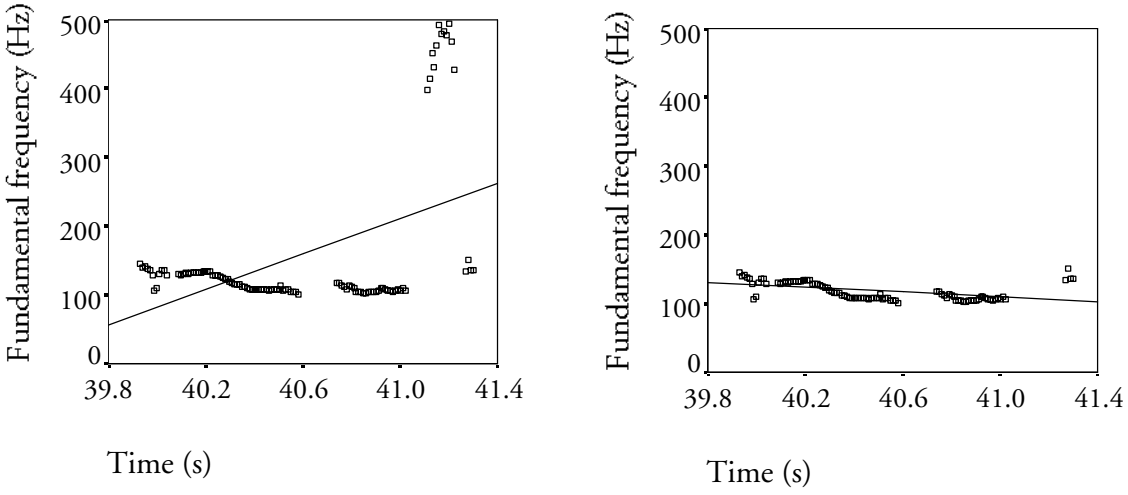


**Figure 4.6** *Scatter plots of the F0 contour of a prosodic phrase and the regression lines calculated using an all-points linear regression technique with and without inclusion of outliers (Nro_ym).*

As shown in Figure 4.7, the variation in F0 slope is far greater in short than long prosodic phrases. This variation reflects the regression analysis' sensitivity to a single

large accent gesture. In short phrases, a single accent fall or rise results in a calculated F0 slope that often differs greatly to that resulting from e.g. a visual abstraction procedure, see Figure 4.8.

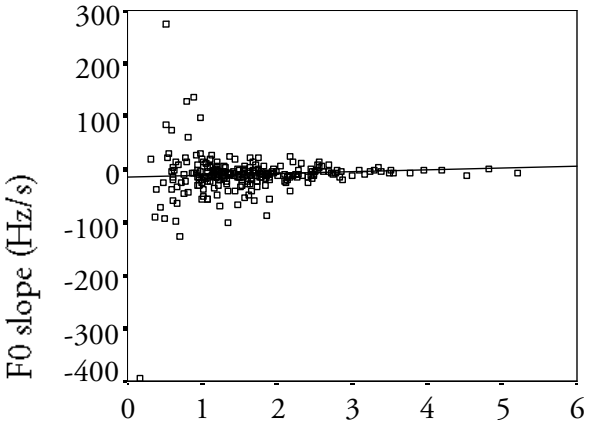


**Figure 4.7** *F0 slope (in Hz/s) as a function of phrase length (in s), for all ten speakers and all prosodic phrases in the transcribed material.*

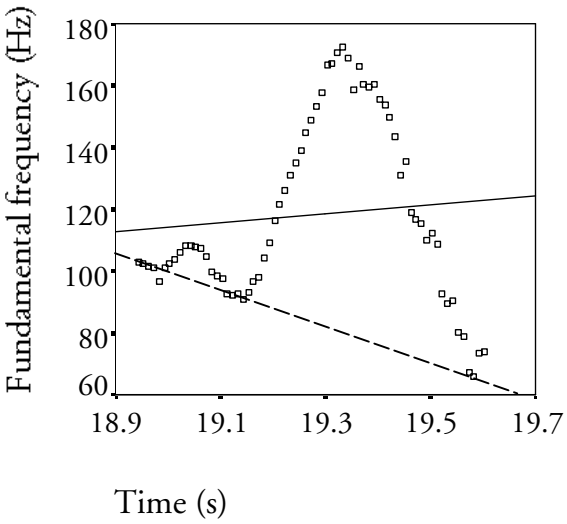


**Figure 4.8** *Scatter plot of the F0 contour of a relatively short prosodic phrase (*och vi hade *'and we had') with only one large accent gesture (the first peak in the F0 contour is a segmental perturbation of F0) (Bar_om). The F0 slope as calculated using a linear regression technique, represented by the full line, is not parallel with the dashed line that represents the result of a visual abstraction procedure. The dashed line has been hand-drawn through the Ls in the phrase.*

In what follows, we will therefore restrict our study to only include the first thirteen[15] phrases longer than one second in each speaker's one-minute fragment, i.e. a total of 130 prosodic phrases, and the measurements taken in these prosodic phrases. Prosodic phrases containing phrase-internal pauses (silent intervals) of more than 200 ms were also excluded since F0 slope is difficult to measure in them as well (regardless of the technique used).

## 4.3 Results

### 4.3.1 Slope and phrase length

The first research question concerned the existence of length-dependent declination in Southern Swedish. In an attempt to try to answer this question, we have calculated the Pearson correlation between F0 slope (in Hz/s) and phrase length (in s) in the data. The graph in Figure 4.9 plots the observed F0 slopes (in Hz/s) as a function of phrase length (in s).

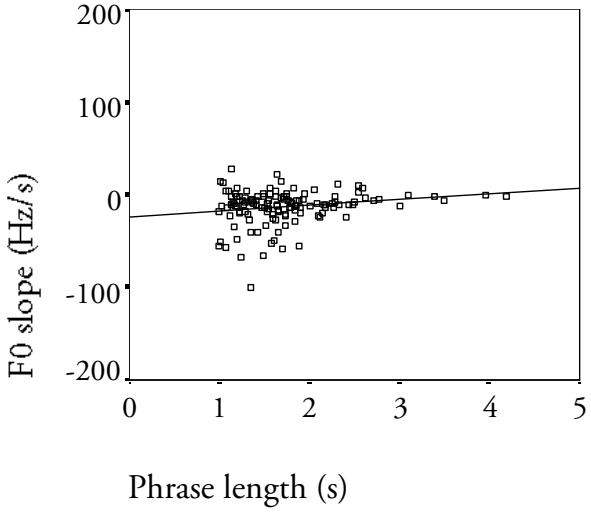


**Figure 4.9** *F0 slope (in Hz/s) as a function of the phrase length (in s), for all ten speakers.*

Although we indeed find a statistically significant correlation at the .05 level between slope and phrase length (R=.20, $R^2$=.04, p<.05), the proportion of the variation in slope accounted for by phrase duration is very small (4%). Based on these results we can reject the null hypothesis, i.e. that there is no correlation between slope and phrase length, but as discussed above, we cannot draw any

[15] The shortest speech section only contained thirteen prosodic phrases longer than one second and without phrase-internal pauses (silent intervals).

conclusions as to whether the correlation between F0 slope and phrase length is due to time-dependent declination or downstep. A weak correlation is expected even in models where all downward trend of F0 is assumed to be due to downstep.

### 4.3.2 Slope and F0 starting point

The second research question concerned the existence of downstep in Southern Swedish. In an attempt to answer this question, we have calculated the Pearson correlation between F0 slope (in Hz/s) and starting point (in Hz) (between F0 slope and start value and between F0 slope and initial accent peak).

The correlation between F0 slope and start value is statistically significant at the .001 level and the correlation is considerably stronger ($R=-.56$, $R^2=.31$, $p<.001$) than the correlation between F0 slope and phrase length. Note that what is relevant here is the relative strengths of the correlations. Since much of the variation in the data can be ascribed to differences between the speakers, whether or not the start value measured in the phrases occurs in an accent peak or not, miscalculated slopes, etc., the exact amount of variation in F0 slope accounted for is not relevant. The graph in Figure 4.10 plots the observed F0 slopes (in Hz/s) as a function of the start values (in Hz).

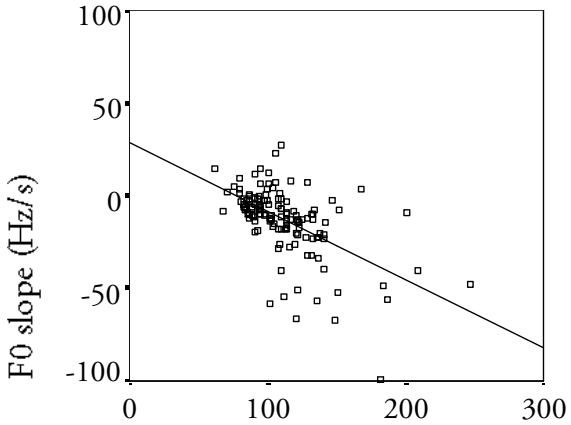


**Figure 4.10** *F0 slope (in Hz/s) as a function of the start value (in Hz), for all ten speakers.*

The graphs in Figure 4.11 plot the observed F0 slopes (in Hz) as a function of the start values (in Hz) for each speaker separately. The relationship demonstrated in Figure 4.10 could have reflected only a tendency for speakers with high starting points to have steeper slopes than speakers with low starting points, rather than a

tendency for each speaker to vary F0 slope with starting point. However, as demonstrated in Figure 4.11, all speakers show a trend toward steeper slopes in phrases with high start values, and more gentle slopes in phrases with low start values.

In the graph illustrating the relationship between F0 slopes and start values in the speech of one of the speakers from Broby (*Bro_om*) one outlier (a slope clearly miscalculated by the linear regression technique in the sense that the calculated slope differs greatly from the slope resulting from a visual abstraction procedure) has been removed.

When using the value of the initial accent peak in each phrase instead of the value in the first stable part of the F0 contour as a measure of the F0 starting point, the correlation between starting point and slope was somewhat weaker ($R=.53$, $R^2=.28$, $p<.001$). The weaker correlation resulting from the initial accent peak's value is most likely due to the peak values' greater sensitivity to differences in degree of prominence between the initial accents in different prosodic phrases (initial accent peak: mean=134 Hz and s.d.=38 Hz, start value: mean=110 Hz and s.d.=26 Hz).

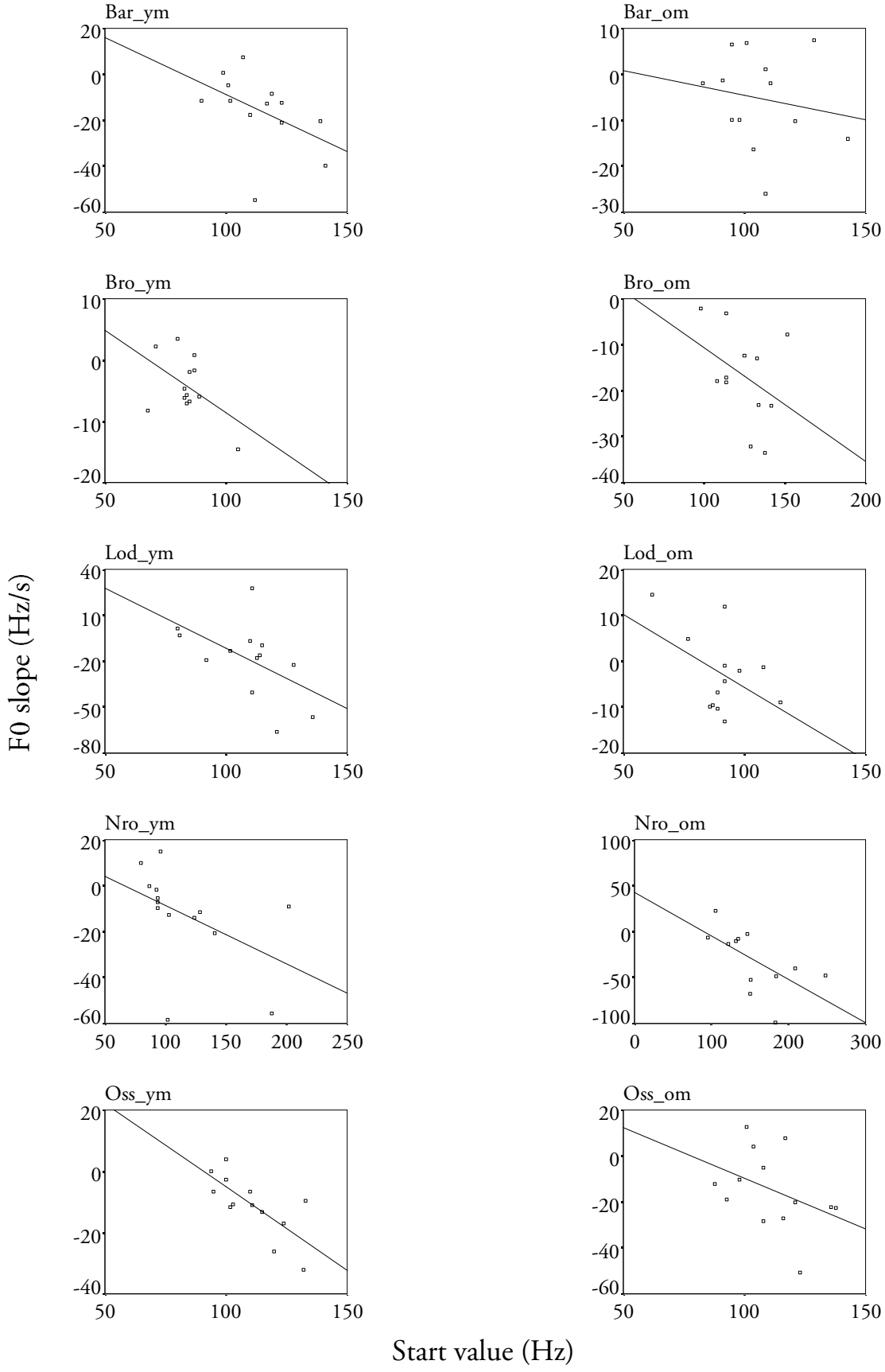


**Figure 4.11** *Slope (in Hz/s) as a function of the start value (in Hz), for each speaker separately.*

### 4.3.3 F0 starting point and phrase length

The third research question concerned the relationship between the speaker's choice of starting point and phrase length. In attempting to determine whether speakers choose higher starting points in long prosodic phrases than in short phrases, we have calculated the correlation between start value and phrase length, as well as between initial accent peak and phrase length. A weak correlation between start value and phrase length ($R=-.19$, $R^2=.03$, $p<.05$) was found. However, the correlation is negative, i.e. the start values show a tendency to decrease with increasing phrase length. Furthermore, the correlation is very weak and the correlation between initial accent peak and phrase length is non-significant ($p>.05$). We therefore conclude that there is no relevant connection between F0 starting point and prosodic phrase length in our data. The graph in Figure 4.12 plots the observed start values as a function of the phrases' length.

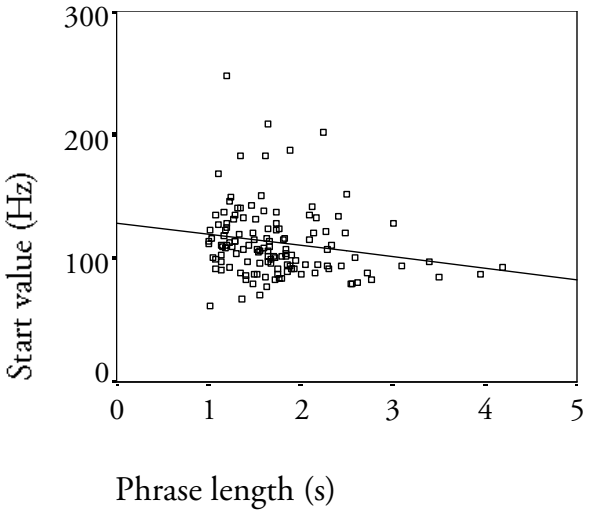


**Figure 4.12** *Start value (in Hz) as a function of phrase length (in s), for all ten speakers.*

## 4.4 Discussion

The downward trend of F0 can be described in various ways: as the result of 1) a global time-dependent downslope, 2) a downscaling of accents with reference to the preceding accent, and 3) final lowering. Several different combinations of the three above-mentioned mechanisms have also been suggested. For example, in Pierrehumbert (1980), the downward trend of F0 in American English is explained as a combination of a peak-by-peak lowering and a global declination line that is fitted to the entire utterance. In Liberman and Pierrehumbert (1984), on the other

hand, the idea of a declining baseline is dropped and it is proposed that the major factors shaping the F0 contour are local ones: a combination of a final lowering effect and the frequent usage of stepping accents. Gussenhoven and Rietveld (1988) claim that the declination effect in Dutch is due to final lowering and time-dependent downsloping. Prieto (1998) and Prieto, Shih, and Nibert (1996) explain the lowering of the H values in Mexican Spanish as the result of downstep and final lowering, but note that the L values are further affected by the distance in time to the preceding peak. The lack of control over factors such as degree of prominence on phrase-initial accents in our data, does not allow us to draw any conclusions as to size and existence of several different contributing factors that affect the scaling of two-tone accents in Southern Swedish.

In the two versions of the Lund model, two opposite approaches are taken to the description of the downward trend of F0: in the original version, the downslope of F0 is explained entirely by time-dependent declination, and in the revised version, the downslope is explained entirely by downstep. The objective of the present study was to answer two questions that will allow us to draw conclusions as to whether our spontaneous speech material is best described by the original or the revised Lund model, and one question that will give us further insights into the degree of preplanning involved in prosodic phrasing of spontaneous speech.

The correlation found between F0 slope and phrase length can be interpreted as evidence to support that at least some part of the downward trend of F0 is due to length-dependent declination. However, the proportion of the variation in F0 slope accounted for by phrase length is very small, and a weak correlation between F0 slope and phrase length was expected even in a model with a peak-by-peak lowering function and no time-dependent down-sloping at all.

The correlation found between F0 slope and F0 starting point is considerably stronger than that which was found between F0 slope and phrase length. We interpret this finding as evidence supporting the revised Lund model.

The lack of a clear relationship between phrase-initial accent peak values and phrase length can be interpreted as further evidence suggesting that the preplanning of prosodic phrases makes use of little or no lookahead. The so-called 'hard preplanning' (Liberman and Pierrehumbert 1984) of whole phrases for which evidence has been found in studies of read speech, is difficult to trace in our spontaneous speech material. There is no evidence to suggest that speakers vary F0 slope to accommodate for differences in phrase length, nor evidence to support the hypothesis that speakers begin a long phrase higher than a short phrase. In the

introductory section, it was mentioned that the two approaches have different implications for the amount of preplanning needed. The revised Lund model, where downstep is modeled, assumes little or no lookahead, and would appear to be the model that best describes our data.

The fact that the downward trend of F0 can be observed in prosodic phrases that clearly have not been entirely planned in advance (see the examples of phrase-internal pauses in section 4.1.1), or that are reorganized as the speaker talks (see (4f)), gives support to the conclusion that no lookahead is required to produce downstepped accents. The utterance *Man hade […] lite höns som hustrun i allmänhet stod för* 'They had […] some hens that the wife usually looked after' is syntactically well-formed. However, with the addition of *hönseriet* 'the hen-house' the relative clause is reorganized into a new sentence: *Hustrun, i allmänhet, stod för hönseriet* 'The wife usually looked after the hen-house'.

(4f)
man hade mjölkdjur | man hade lite grisar | och lite suggor | och lite höns | som hustrun i allmänhet stod för hönseriet | (*Lod_om*)

'they had milk cows | they had a few pigs | and a few sows | and some hens | that the wife usually looked after the hen-house |' (*Lod_om*)

The lack of clear evidence supporting the hypothesis that the starting points vary in order to accommodate for different phrase lengths nevertheless raises an interesting new question. Is the variation observed in starting point values an expression of discourse prosody rather than a feature of the planning and realization of the prosodic phrase? Do the starting points vary in a systematic way over the course of larger units of speech? Is there a prosodic structuring of speech into units larger than the prosodic phrase, and in that case, what implications does that have for our conception about how much preplanning is involved in spontaneous speech?

### 4.4.1 Observations on the qualitative behavior of downstepping accents in Southern Swedish

After having determined that the downward trend of F0 in our spontaneous speech material is better described by the revised than the original Lund model, we would like to address a few questions regarding the qualitative behavior of downstep in Southern Swedish. One question, mentioned already in the introductory section of this chapter, is particularly interesting, namely the issue of the downstepping behavior of word accent II in Southern Swedish. In the following, we will present a

discussion that is based on observations made during the process of carrying out the investigation presented above[16]. It should only be seen as a first attempt to an analysis of the behavior of Southern Swedish word accents in a larger perspective, and as a base for further studies and discussions.

#### 4.4.1.1 Triggering of downstep

To our knowledge, only one study of downstep in Southern Swedish has been undertaken, and the focus of that study was on tonal coupling between prosodic phrases (Bruce 1984). However, differences between Stockholm Swedish and Danish as regards downstep have been reported (Thorsen 1983), and since Southern Swedish shares many prosodic properties with Danish (Gårding *et al.* 1974), it is reasonable to assume that some differences may also exist between Stockholm Swedish and Southern Swedish.

One characteristic of downstep in Stockholm Swedish is that it is trigged by focus (Bruce 1982a), i.e. downstep mainly occurs after focus (the H phrase accent) in the utterance. In Danish, as in Southern Swedish, there is no (high) phrase accent, and downstep occurs even before focus (Thorsen 1983). This is also the case in Southern Swedish where downstep can occur pre-focally as shown in Figure 4.13 (an example of post-focal downstep is given below in Figure 4.14). In the tonal transcriptions given in the middle tiers in the figures, no difference is made between focal and non-focal accents since the only difference between these accents in Southern Swedish is related to the accent's pitch range. The focal accents are, nevertheless, indicated with ‘ˈˈ’ and non-focal accents with ‘ˈ’ in the orthographic transcription (in the top label tier of each figure).

[16] Some phrases produced by the female speakers representing *Skåne* in *SweDia*'s public database are also discussed.

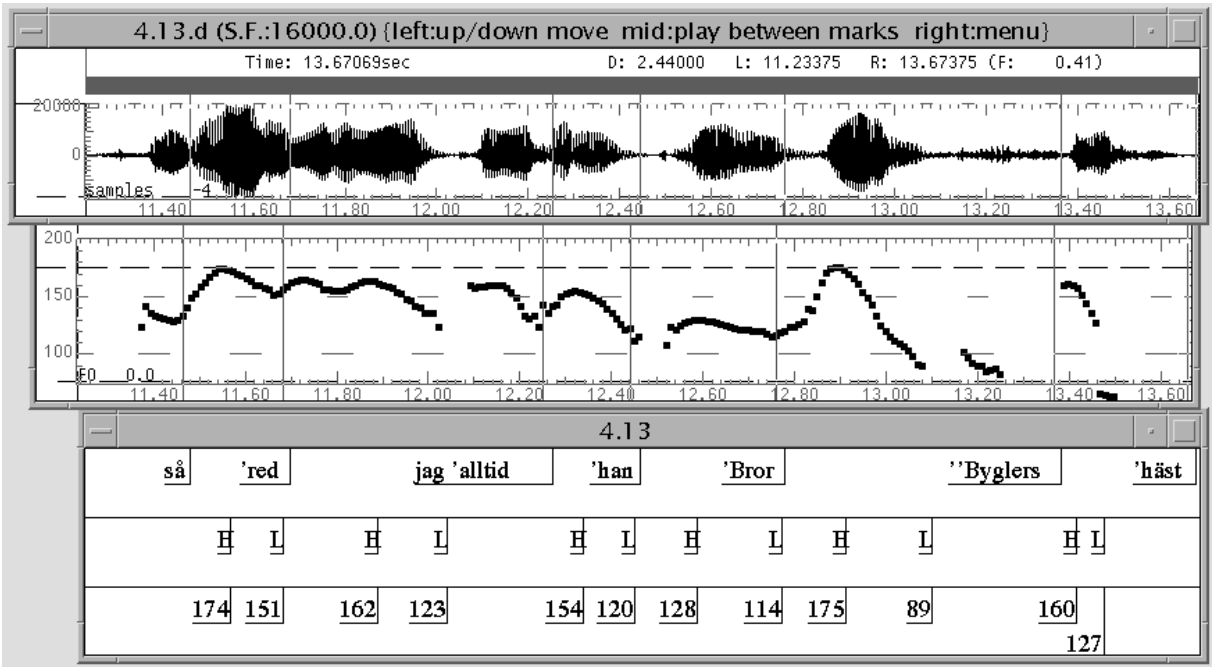


**Figure 4.13** *Speech wave and F0 contour of the prosodic phrase* så red jag ju alltid han Bror Byglers häst *'I always rode his – Bror Bygler's – horse' (Bro_om). The three label tiers are, from top to bottom: 1) a word tier, 2) a tone tier with the accent H and L tones indicated, and 3) a tier where the F0 values of the H and L tones are given in Hertz. Each H and L tone (in Hz) of the non-focal accents is lower than the preceding H and L tone, respectively.*

#### 4.4.1.2 L scaling

The L tones' scaling has not been given the same degree of attention as the scaling of H tones in the literature. One interesting question as regards the scaling of L tones is how they behave under prominence. Van den Berg, Gussenhoven and Rietveld (1992: 347) note that while "the effect of increased prominence on H* would […] appear to be uncontroversial, the effect of increased prominence on L* is less clear". Liberman and Pierrehumbert (1984) suggest that the scaling of L tones is symmetric to that of H tones. They assume that the transformed value of a L is the negative of the transformed value of an equally prominent H. They leave the question as to how two-tone accents behave under changes in prominence unanswered, however. It is exactly the two-tone case that is relevant here, and whether or not the two tones behave in the same manner (e.g. if the downstep constant is the same for the two tones) under changes in prominence as well as under equal prominence relations.

In the work undertaken by Bruce on downstep, both tones of the word accents are assumed to be subject to downstep (but converge to different values). Only the

tones in accent II (H*L in Standard Swedish) words were examined, since the phonetic difference between accent I and II is assumed to be one of timing, and the results for one accent are assumed to be easily applied to the other accent. However, it is not necessarily the case that the L tones behave in the same way in accent I words as they do in accent II words, as has been demonstrated by Fant and Kruckenberg (1994). Nor do we know if the results for HL accents can be applied to the Southern Swedish accent II, an accent that may be analyzed as a LH accent. We will therefore discuss L scaling by examining word accent II. How L tones behave under prominence is discussed in section 4.4.1.3.

Like in all dialects of Swedish (except Finland Swedish), there are two types of accents in Southern Swedish: accent I (the acute word accent) and accent II (the grave word accent) (Bruce and Gårding 1978). Within a ToBI framework (Beckman and Ayers 1993, Silverman *et al.* 1992), accent I can be transcribed as a H*L accent (like the Stockholm Swedish accent II), since a peak is found at the beginning of the stressed syllable after which F0 falls throughout the remaining part of the syllable. Word accent II is also perceived as an accent that ends low, but the accent fall starts at the end of the stressed syllable, and is preceded by a rise that starts at the beginning of the stressed syllable. Within the ToBI system (as well as in the ToBI based tonal transcription system for Swedish) where tri-tonal accents are avoided, the Southern Swedish accent II may be transcribed as a bi-tonal L*H accent, like the Danish accent (Thorsen 1983). However, a description of the accent as a falling one ending low is motivated by the fact that sequences of varied accent types demonstrate a pattern where a low turning point rather than a F0 plateau is often found between accent II and accent I, i.e. between the H of the hitherto called L*H accent (accent II) and a following H*L accent (accent I), even when very few unstressed syllables intervene, see Figure 4.14. In this example, only one unstressed syllable intervenes between the stressed syllable of the accent II word *hittar* 'find' and the accent I word *dom* 'they'. Nevertheless, a low turning point[17] can be clearly observed indicating the presence of a L tone. In the orthographic transcription tier, accent II words are transcribed with a '`' above the stressed vowel. The first turning point of word accent II is furthermore labeled and placed within parentheses in the tone tier in the figures below.

[17] Note that we are not using the terms 'tone' and 'turning point' synonymously. A turning point is not necessarily a reflex of a tone, but simply a phonetic turning point accessible to empirical observation in the F0 contour.

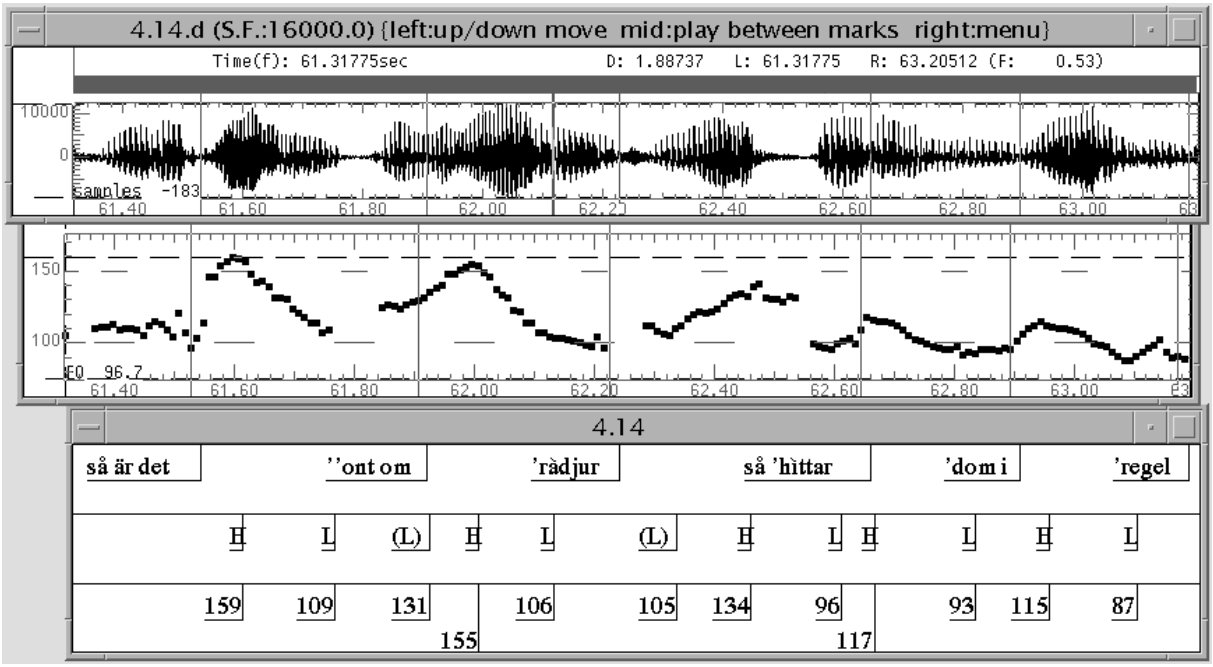


**Figure 4.14** *Speech wave and F0 contour of the prosodic* phrase så är det ont om rådjur så hittar dom i regel *'so if there is a shortage of roe deers then they usually find them' (Oss_ym).*

One way in which we possibly could test whether it is the low turning point preceding or following the H tone that is lexically specified as a tone is given in Ladd (1996). Ladd discusses the results of the previously mentioned experiment undertaken by Fant and Kruckenberg (1994), in which it is shown that the L tone of word accent II in standard Swedish (H*L accent) is not the same as the L tone of word accent I (HL* accent). In Bruce's (1977) analysis of the standard Swedish word accent distinction, both word accents are described as falls, HL, and the phrase accent as a H tone. The only difference between accent I and accent II is that accent II has a later alignment of the word accent fall than accent I. However, Fant and Kruckenberg (1994) show that when speakers pronounce the word accents with varying degrees of emphasis, the L tone of accent II behaves more 'real' in the sense that it is lowered with increased emphasis, whereas in accent I, the L increases in F0, just as the two H tones do.

Figure 4.15 demonstrates the effect that the degree of emphasis has on the L tone in a Southern Swedish accent I word (H*L) as compared to the two low turning points of an accent II word. A male speaker of Southern Swedish was asked to read the sentences *Dom målen mellan* 'Those meals between' and *Dom mellan målen* 'Those between the meals' with varying degrees of emphasis on the words *målen* and *mellan* (three versions with different degrees of emphasis on the phrase-internal word and one non-focally accented version where focus was placed on the phrase-

final word). The phrase-internal words *målen* (accent I) and *mellan* (accent II) were subsequently excised from their context, although the preceding and following nasals were kept in order to make sure that the low turning points be easily observable in the F0 tracings. In the case of *målen* (which has an earlier timing of the word accent fall than the accent II word *mellan*) the entire phrase-initial word *dom* was kept in order to make sure that the low turning point preceding the accent fall could be seen. In both words' F0 contours, three turning points can be observed: a low turning point followed by a rise to a high turning point which in its turn is followed by a low turning point. In the accent I word *målen* 'the meals' the phonological analysis is unproblematic; it is the second and third turning points that are interpreted as reflexes of the two tones H and L.

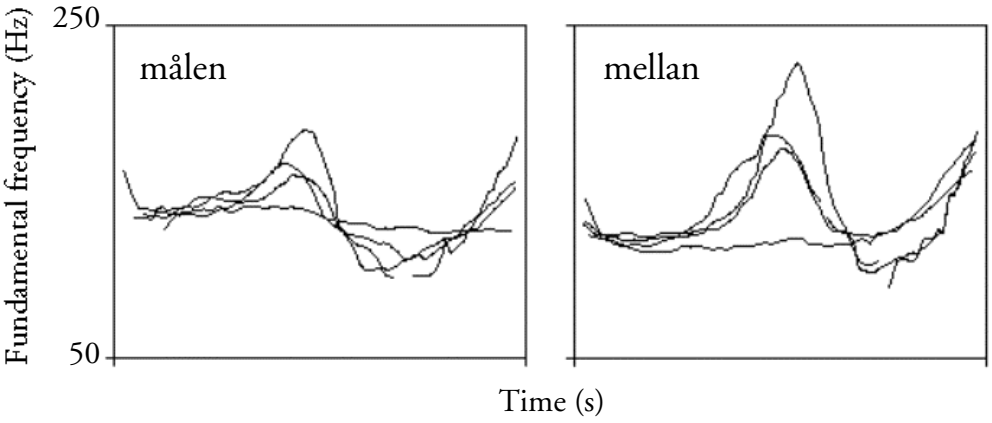


**Figure 4.15** *F0 contours of the accent I word* målen *'the meals' and the accent II word* mellan *'between' pronounced with varying degrees of prominence (male Southern Swedish speaker).*

It is clear that it is the second low turning point in the accent II word that behaves like the L tone in the accent I word, i.e. by decreasing in F0 with increased degree of emphasis. The low turning points preceding the accent falls are unaffected by differences in degree of emphasis in both words. We interpret this finding as further evidence suggesting that it is the second low turning point of accent II that is lexically specified (although it does not rule out an analysis in which the word accent II L*H is analyzed as followed by a L phrase accent in focal position).

The reason why a L*H analysis of the Southern Swedish accent II is unsatisfactory becomes evident when we examine how accent II is downstepped in the data. In sequences where an accent II (the hitherto called L*H accent) follows an accent I (H*L), the first low turning point of the second accent is not downstepped in relation to the preceding accent's L tone. Instead it is the L turning point after the H tone in the accent II that would appear to be downstepped in relation to the preceding accent's L tone, see Figure 4.16 (and Figure 4.14 above). This pattern

makes sense only if we analyze the accent II as a two-tone accent where the first tone is high and the second low. This analysis is also consistent with Bruce and Gårding's (1978) analysis of the Swedish word accents, where the word accents are described as falls with different synchronization with the stressed syllable.

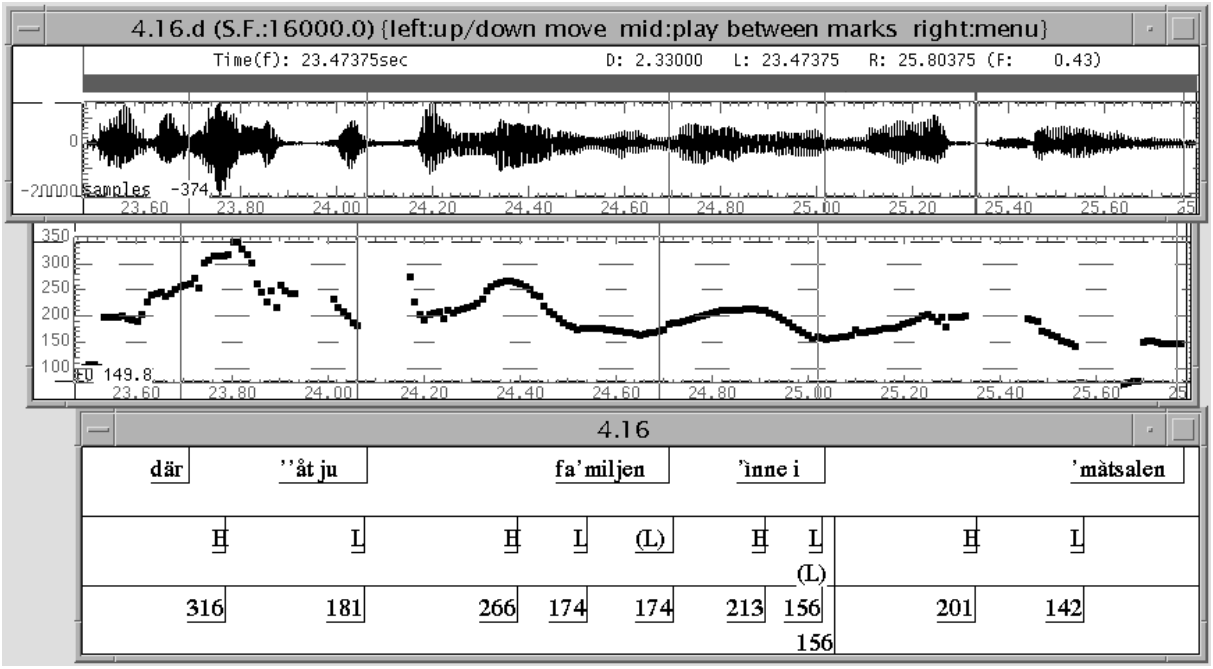


**Figure 4.16** *Speech wave and F0 contour of the prosodic phrase* där åt ju familjen inne i matsalen *'there the family ate in the dining-room' (Bro_ow).*

### 4.4.1.3 Prominence relations in downstepped sequences of accents

An interesting question as regards downstep in Southern Swedish that needs further investigation concerns focal accentuation. When focus is found in non-initial position, it is clear from our observations that the focal accent differs from the expected downstepping pattern in that it is not produced with a lower accent peak than the preceding accent. The focal accent has a higher peak than expected, and often even a lower valley than expected given the relevant step size, which is consistent with the tendencies observed in the limited read material presented in the previous section, see Figure 4.15 above. It is also clear that the (non-focal) accent following the focal accent has a lower peak than the preceding accent.

What is not so clear from our observations of non-controlled spontaneous speech, is the L tones' behavior under different degrees of prominence in downstepping sequences. In Figure 4.17, the focally accented word has a higher peak than the preceding non-focally accented word, but its valley is not lower than the preceding accent minimum, possibly indicating that no downstepping has occurred between

the phrase-initial and the following accent[18]. After the focal accent, on the other hand, we see a clear downstepped sequence of accents where each accent is downstepped in relation to the preceding one. The lack of a clear difference in the L tones' values between the focally accented word *åt* and the following non-focally accented instance of *åt*, indicates that prominence relations are taken into account even in the downstepping of L tones. In Pierrehumbert (1980: 79), prominence relations are taken into account in the following way:

(4g) $$H^*_{i+1} = H^*_i \ * \frac{\text{Prominence } (H^*_{i+1})}{\text{Prominence } (H^*_i)}$$

In a sequence of accents, the most prominent accent does not only have a higher peak than the preceding accent (or in a downstepped sequence of accents, a higher peak than expected given the relevant step size), the accent following the prominent accent also has a lower peak than the preceding prominent accent (or lower than expected in a downstepped sequence of accents). In Liberman and Pierrehumbert (1984), it is suggested that the scaling of L tones is symmetric to that of H tones, meaning that the most prominent accent does not only have a lower valley than the preceding accent (or in a downstepped sequence of accents, a lower value than expected given the relevant step size), the accent following the prominent accent also has a higher valley than the prominent accent (or higher than expected in a downstepped sequence of accents). Figure 4.17 demonstrates an example of this type of L tone behavior. Although there is strong evidence of downstepping occurring between the two instances of *åt*, the second and less prominent instance has only a negligibly lower valley than the preceding, more prominent accent's valley.

[18] Even assuming that no downstep has occurred, it is surprising that the valley of the focally accented word is not lower than the preceding, non-focally accented word's valley.

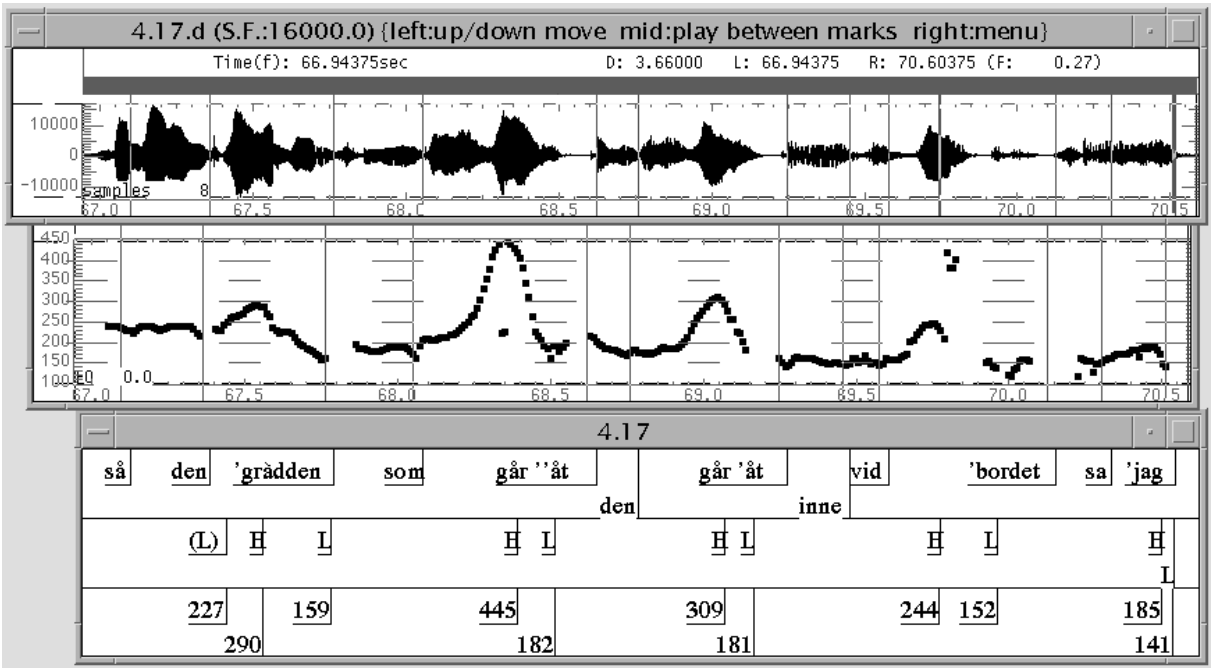


**Figure 4.17** *Speech wave and F0 contour of the prosodic* phrase så den grädden som går åt den går åt inne vid bordet sa jag *'so the cream that is used is used at the table I said' (Bro_ow).*

On the other hand, the same cannot be said for the example given in Figure 4.16 above where the difference in F0 between the focally accented word *åt* and the following non-focally accented *familjen* is obvious in the measurements of both tones (despite the fact that the inherent F0 of the vowel in which the second peak occurs, /ɪ/, is high). Furthermore, it is not evident that the difference in peak height between a focally accented word and a following non-focally word is larger than that between two successive non-focally accented words in downstepped sequences of accents. In Figure 4.17 above, the largest step size is found between the focally accented *åt* and the following non-focally accented *åt*, but a large step size would also have been expected if both words were non-focally accented given their early position in the downstepped sequence of accents. Assuming that prominence relations are taken into account as modeled in (4g), the non-focally accented instance of *åt* should have the same or (if downstepping is assumed to have occurred at least between the two instances of *åt*) a lower peak than *grädden*. It does not, however, and it would rather seem as if the second instance of *åt* is downstepped in relation to the preceding instance of *åt*, without taking into account the difference in degree of prominence.

It is possible that prominence relations between accents are signaled in a different manner in spontaneous speech than has previously been observed in read, laboratory speech. Preliminary observations of our Southern Swedish data seem to

suggest that the most important way for signaling prominence relations is in the relation between the more prominent (focally accented) accent and the preceding accent rather than in the relation between the prominent accent and *both* the preceding and following accent, at least when it comes to H scaling[19].

The analysis of the L tones is more problematic, since the L tones in our data demonstrate a less systematic behavior than the H tones. On the other hand, the L tones' exact F0 values are likely less important perceptually, and perhaps therefore simply not as carefully targeted as the H tones' values. Alternatively, it may prove fruitful to look for a relation between the L tones' values and the value of the preceding H tone of the same accent (rather than the value of the preceding accent's L tone)[20]. Further studies and more highly controlled experiments are needed here, preferably using elicited spontaneous speech.

### 4.4.2 Summary

In the present chapter, we have investigated the existence and, to some extent, the nature of downstep in a spontaneous Southern Swedish speech material. By examining the relationships between F0 slope and phrase length and between F0 slope and starting point, we were able to determine that the revised Lund model describes our data in a more accurate way than does the original Lund model. The evidence supporting the revised Lund model led us to conclude that very little lookahead is made use of in the planning of prosodic phrasing in spontaneous speech. A starting value is chosen, based most likely on expressive and pragmatic factors (and not on the length of the upcoming prosodic phrase), and thereafter the phonetic values of the accents' peaks and valleys are chosen taking only the previous accent's values into consideration. As noted by Liberman and Pierrehumbert (1984: 220) there is nothing to stop speakers from choosing a higher starting point in longer phrases, but we have not found any evidence to support that such an adjustment would be necessary. That expressive factors affect the F0 starting point, and thereby the F0 register, is known from a previous study of Swedish (Bruce 1982a).

In our registrations of F0 starting points, we measured the F0 value in the phrase-initial accent peak. We did not make any measurements of F0 in the first accent

[19] In other words, the downstep rule appears to apply without taking prominence relations into account both when the accent subjected to the rule has the same degree and a lesser degree of prominence than the preceding accent.

[20] Such a dependency would mean that the L tones are only indirectly subjected to downstep.

valley, although Bruce's model (1982a) of downstep takes the accent minima scaling as a base for the subsequent modeling of the accent maxima (and not vice versa as in e.g. Prieto's model (1998)). The reason for avoiding measuring F0 in the valleys was the observation that the L tone of accent II (if analyzed as a L*H accent) behaved differently from the L tone of accent I (H*L). It would appear that it is the low turning point following the high tone of both word accents that is downstepped in relation to the preceding accent's L. We therefore proposed an analysis of the word accents in which both are described as falling two-tone accents (HL). Our analysis is further motivated by the fact that 1) a low turning point almost always appears between the accents in sequences where an accent II is followed by an accent I, and that 2) the second low turning point of accent II is the turning point that behaves most like the L tone of accent I under changes of prominence. Both tones of the two-tone accents are in some way (directly or indirectly) subject to downstep.

Finally, our observations of how the two-tone accents behave under changes in prominence, led us to suggest that more focus be placed on H scaling in future studies of downstep in Southern Swedish. The H tones demonstrate more consequent downstepping patterns than the L tones, indicating that the L scaling is not an appropriate base for the modeling of the H tones. We have also discussed the possibility that only increases in degree of prominence between successive accents are marked in downstepping patterns by an unexpected high peak and an unexpected low valley (unexpected given the relevant step sizes).

CHAPTER 5

# Tonal coherence among prosodic phrases

## 5.1 Introduction

In the present chapter, we follow up on findings concerning the downward trend of F0 within the prosodic phrase presented in chapter four. It was found that speakers do not adjust the F0 starting point in order to accommodate for the upcoming phrase's length. In other words, speakers do not start longer prosodic phrases higher than short phrases in order to ensure that F0 does not drop below the speaker's F0 floor, or to ensure that some constant or minimum F0 slope can be produced. Nevertheless, the measurements of the F0 starting point – the F0 value in the first stable part of each prosodic phrase's contour and in the phrase-initial accent peak – revealed a substantial amount of variation. Given that this variation cannot be accounted for by phrase length, i.e. by a feature of the prosodic phrase, it seems reasonable to look for an explanation in the discourse. The present chapter is dedicated to investigating tonal coherence among prosodic phrases.

In our usage, the term 'discourse' refers to naturally occurring connected speech. In analyses of discourse, the object under investigation is usually the organization of language above the sentence or above the clause (Stubbs 1983). In analyzing

discourse prosody, we take the relevant organization of language to be the one above the prosodic phrase.

### 5.1.1 Phrasal downstep and tonal coupling

Phrasal downstep, downward scaling of a prosodic phrase's pitch range or register relative to the preceding phrase, provides a possible explanation to some of the variation in F0 starting points our data. The idea of phrasal downstep, a downstepping of the first accent in each phrase relative to the first accent in the preceding phrase, has been put forward by van den Berg *et al.* (1992). In Figure 5.1, the three phrase-initial accents (258 Hz, 241 Hz and 200 Hz, respectively) can be seen as a sequence of phrase-initial accents downstepped relative to the preceding.

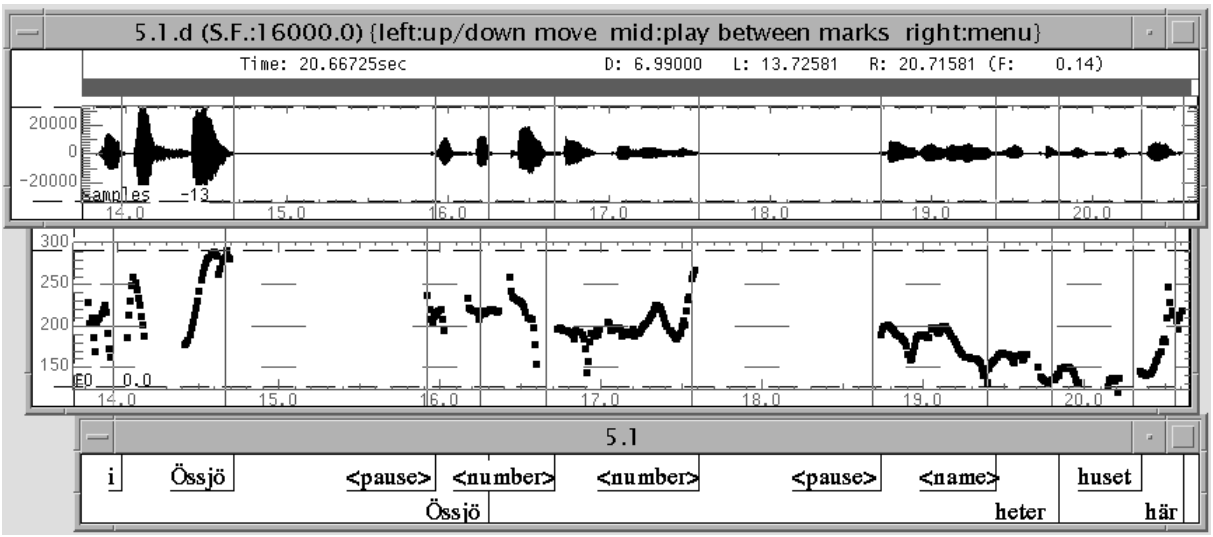


**Figure 5.1** *Speech wave and F0 contour of the three prosodic phrases* i Össjö *'in Össjö',* Össjö <number>[1] <number> *'Össjö <number> <number>' and* <name> heter huset här *'<name> is the house here called (example from the spontaneous recording of Oss_ow).*

In van den Berg *et al.*, a distinction is proposed between accentual downstep (downward scaling of H* targets within the phrase) and phrasal downstep (downward scaling of a prosodic phrase's register relative to the preceding phrase). Phrasal downstep causes the first H*s of the phrases in a sequence of phrases[2] to form a downstepping pattern. Thus the interruption of downstep between prosodic phrases, the F0 reset, is seen not as the result of an upward shift relative to the

[1] In order to guarantee the speaker's anonymity, numbers and names have been omitted.

[2] In van den Berg *et al.*, phrasal downstep was examined in text units of four prosodic phrases (or 'association domains').

preceding accent, but as a downward shift relative to the preceding phrase. In such a phrase-relative model, one expects unambiguously downstepping series of resets and therefore systematic variation in the phrase-initial accent peaks' F0 values.

The idea of something like phrasal downstep also exists in intonation models where the downward trend of F0 is described as the result of time-dependent declination. Cooper and Sorensen (1981), e.g., report on 'clausal declination' that can be observed to be reset at the beginning of each new clause without resetting of the more global so-called 'utterance declination'. Grønnum Thorsen (1988: 16) also describes an intonational organization above the sentence level when she demonstrates that "each sentence is associated with its own declining sentence intonation contour, but together two or three such contours describe an overall downward trend", a gradual decrease through the text (by 'text' is meant "a sequence of semantically but not necessarily syntactically coordinated sentences" (Grønnum Thorsen 1988: 15)).

In Bruce (1982b), the downward trend of F0 over the course of the 'text unit', as Bruce terms it, is described as the result of 'tonal coupling'. Tonal coupling between the utterances (short one-prosodic-phrase-long utterances) within the text unit manifests itself in Southern Swedish in such a way that the earlier part of an utterance's F0 contour "fits into" the later part of the F0 contour of the preceding utterance. The F0 minimum of the first accent in an utterance has a value similar to that of the F0 minimum preceding the last accent in the preceding utterance.

The tonal coupling found in Bruce (1982b), raises an interesting question regarding the function of the starting points' sensitivity to utterance length. In Bruce (1982a), the observed raising of the starting points (first accent's maximum[3]) in proportion to the utterance's length (measured in terms of the number of upcoming accents), is interpreted as a way for the speaker to signal utterance length. However, if the earlier part of a utterance's F0 contour simply is a "copy" of the later part of the F0 contour of the preceding utterance, then its F0 starting point says nothing about its length. That is, the variation in F0 starting points is not dependent on any feature of the utterance, but on the utterance's placement in a larger text unit. The raising of the text unit's starting point depends not on the first utterance's length, but on the length of the entire text unit. Indeed, Bruce (1982b) also shows that the F0 peaks and valleys of a single-utterance text unit are lower than those of the initial utterance in a larger unit.

[3] Note, however, that in Stockholm Swedish, the first accent maximum is not necessarily the first phrase accent's maximum (i.e. the F0 point that is claimed to vary with utterance length).

There is one important difference between the approach of van den Berg *et al.* (1992) and Bruce (1982b). In Bruce's approach, the entire text unit forms one declination raft or sequence of stepping accents, interrupted only by the last accent minimum in each utterance which falls below the raft, and the first accent minimum in each (non-initial) utterance which is a copy of the penultimate accent minimum in the preceding utterance. In the approach of van den Berg *et al.*, on the other hand, a new sequence of stepping accents is thought to begin in each prosodic phrase and its F0 starting point depends on the starting point of the preceding sequence's starting point rather than on its F0 value in the end point. It has been argued that a similar pattern might also have been found in the Swedish data examined by Bruce (1982b) if the number of accents in the utterances had been varied instead of kept constant.

Here, we will not be able to draw any conclusions as to whether it is phrasal downstep or tonal coupling that best describes our data. What we are interested in knowing, is whether the decrease in F0 over the course of several prosodic phrases – either as the result of phrasal downstep or tonal coupling – is used to signal coherence in spontaneous speech. In both the phrasal downstep and tonal coupling approach, the tonal signaling of coherence among prosodic phrases is thought of as a lowering of F0 starting points in successive prosodic phrases, rather than as a single local signal (e.g. a high accent peak at the beginning of the coherent group of prosodic phrases and/or an abrupt decrease in F0 at the end). In the present chapter, we will look for signs of tonal coherence signaling among prosodic phrases by examining the decrease in F0 across several successive prosodic phrases.

Note that we do not expect phrasal downstep to account for all the variation in F0 starting points our data. We know from previous studies that the F0 starting point, at least when measured in the phrase-initial accent peak, is also subject to factors such as the degree of prominence assigned to the phrase-initial word. The degree of prominence that the speaker chooses to assign to the phrase-initial word is influenced by factors such as the word's information status, e.g. whether it is 'new' or 'given' in the discourse, and whether it occurs in a broad focus or as the single element in a narrow or contrastive focus. Expressive factors such as the degree of involvement or excitement have also been demonstrated to affect the F0 register chosen by the speaker and the height of the phrase-initial accent peak (Bruce 1982a, Liberman and Pierrehumbert 1984). The example in Figure 5.2 illustrates two prosodically coherent prosodic phrases where phrasal downstep/tonal coupling cannot be observed. Although the two phrases are clearly coherent semantically, and in several aspects also prosodically (there is e.g. no silent pause between the

phrases), the phrase-initial accent peak in the second phrase is higher than in the first phrase.

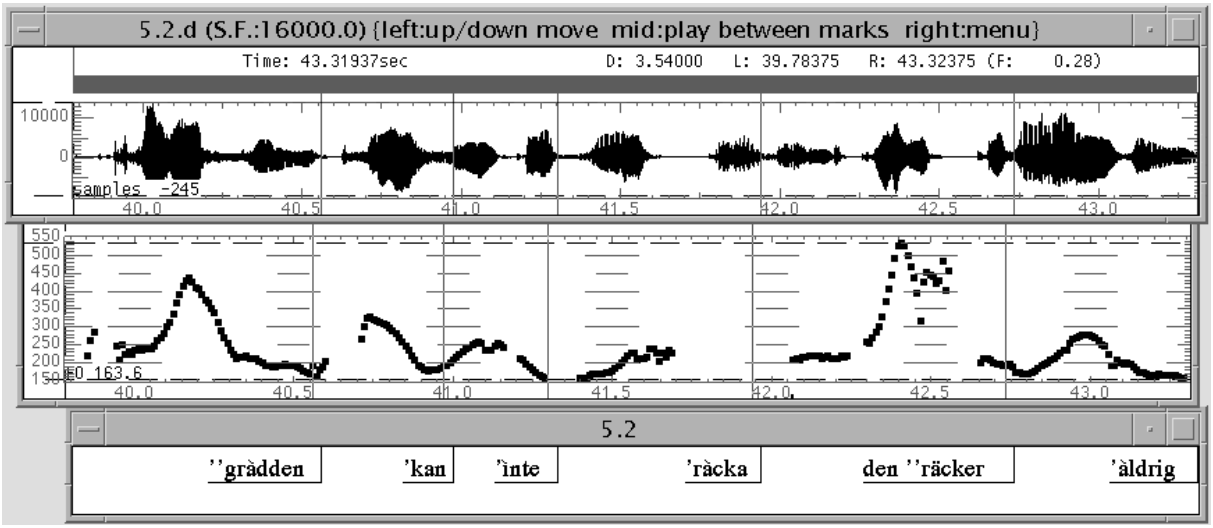


**Figure 5.2** *Speech wave and F0 contour of the two prosodic phrases* grädden kan inte räcka *'the cream cannot be enough' and* den räcker aldrig *'it is never enough' (Bro_ow).*

### 5.1.2 Research question

The research question to be addressed in this chapter is whether F0 is used in spontaneous speech to signal coherence among prosodic phrases. To answer this question, we need to investigate both acoustic aspects of prosodic phrases in spontaneous speech and the perception of prosodic phrase boundaries between and within the units where phrasal downstep or tonal coupling operates.

## 5.2 Method

### 5.2.1 Speech material

The speech material investigated was described in section 4.2.1. It is the speech of ten male subjects who represent the region *Skåne* in *SweDia 2000*'s public database. The speech of five female subjects (from the older (*ow*) and younger generation (*yw*)) representing the same region has furthermore been included (*Bar_yw*, *Bro_ow*, *Bro_yw*, *Oss_ow* and *Oss_yw*).

The speech sections are typically short anecdotes dealing with the older subjects' youth and descriptions of the young subjects' work. They are clearly longer than what is generally considered to be a speech paragraph or text unit (Bruce 1982b).

### 5.2.2 Procedure and measurements

We attempted to get an insight into the tonal organization of prosodic phrases in discourse by using several methods.

Firstly, we recorded two different measures that can be expected to vary over the course of the larger speech unit to give the impression of tonal coherence among prosodic phrases: the prosodic phrase-initial accent peak and the start value (measured in the first stable part of the prosodic phrase's F0 contour). We refer to these two measures as measures of the F0 starting point. The F0 start value was used in combination with the measure of the phrase-initial accent peak because it was expected to vary less with the degree of prominence assigned to the phrase-initial word than the F0 measure of initial accent peaks. Each prosodic phrase's position in the audio file was also recorded.

Prior to the labeling of the F0 starting points, a prosodic segmentation into prosodic phrases was done interactively using the speech analysis program ESPS/Waves+™. Since the ten male subjects' speech was segmented and labeled in the previous study on F0 downtrend (see chapter four), only the female subjects' speech had to be segmented and labeled at this stage. The labeling of prosodic phrase boundaries (in the male subjects' speech) was evaluated by comparing it to the labeling made by two more experienced transcribers (see section 4.2.2).

Using the ESPS/Waves+™ *get f0* function, F0 tracings were generated for the female subjects' audio files. Working interactively, listening to the speech and observing the F0 tracings, two starting points (as described above) in each prosodic phrase's F0 contour were then marked in a label tier. The F0 values of the F0 starting points were extracted automatically using the labels placed in the label tier.

Since little is known about prosodic phrasing and discourse prosody in spontaneous Swedish, the investigation had an exploratory nature and a number of different methods were employed in order to answer the research question at hand. In the first stage of the study, we used the measurements of F0 starting points to group the prosodic phrases in the material into sequences of downstepped/tonally coupled phrases. Secondly, we looked for evidence of perceived tonal coherence within groups of prosodic phrases formed on the basis of F0 starting points. Without coherence being perceived within the downstepped/tonally coupled phrases, we cannot make any claims as to the coherence signaling function of decreasing F0 starting points. We used the labeling of boundary strength made by the expert transcribers (see section 4.2.2) to determine whether boundaries between prosodic

phrases contained in the same sequence of downstepped/tonally coupled phrases in fact are perceived as weak and boundaries between prosodic phrases pertaining to different sequences as strong. Thirdly, we touch on the subject of where strong and weak boundaries respectively are produced and/or perceived in relation to linguistic structure. Whether perceived or not, tonal coherence signaling is only interesting to us if it has a function in communication; if it e.g. serves to make the understanding of speech easier by grouping prosodic phrases that are semantically related. Finally, we examined disfluent parts in the material (phrases containing phrase-internal pauses, disfluencies (speech repairs) and signs of syntactic reorganization) in order to get some insight into the amount of preplanning needed to produce tonally coherent sequences of prosodic phrases. A prerequisite for the role of declining F0 in coherence signaling in spontaneous speech, is that it can be produced with no or only a limited amount of preplanning.

## 5.3 Results and discussion

### 5.3.1 The domain of phrasal downstep

As a first step to approach possible tonal coherence signaling among prosodic phrases in our data, we needed to make sure that the speech units under investigation (the approximately one-minute long sections) were larger than the unit within which lowering of F0 starting points across successive prosodic phrases (phrasal downstep or tonal coupling) operates. The literature on the subject of declination and downstep indicates that there may be at least two separate domains of declination/downstep: the prosodic phrase (clausal declination or accentual downstep) and the prosodic utterance or speech paragraph (utterance/paragraph declination or phrasal downstep). Although the speech units investigated were about one minute long and contained between 15 and 31 prosodic phrases, they could constitute a speech paragraph, or even only a part of such a larger domain.

We established possible relationships between a phrase's position in the speech unit and its start value and phrase-initial accent peak respectively, by applying the Spearman Rank-Order Correlation (Rho).

Much to our surprise, even the analysis of such large speech units as those investigated, revealed some cases of F0 decrease across all the prosodic phrases contained in the speech unit (observed as a negative correlation between initial accent peak and position, and/or between start value and position). A significant negative correlation could be observed in four of the subjects' speech ($p<.05$),

although the correlation could only be observed in both measurements of F0 starting points in one of the speech units. In one speech unit, a significant positive correlation was found between start value and position ($p<.05$). The fact that the average decrease in F0 between successive phrase-initial accent peaks and between successive start values was very small, made it easy to question its perceptual relevance, as did the fact that a positive correlation was found in one of the speech units. Furthermore, the negative correlations between F0 starting points and position – where such could be observed – were not the result of uninterrupted decreasing F0 starting points. It would be misleading to interpret these results as indicating the presence of uninterrupted phrasal downstep, or as indicating the presence of an uninterrupted declination ramp stretching over the entire speech unit, as demonstrated in Figure 5.3. In this speech unit (containing 21 phrases), negative correlations between both initial accent peak and position ($\rho=-.63$, $p<.05$) and between start value and position ($\rho=-.46$, $p<.05$) were found. The prosodic groups of interest, prosodic phrases demonstrating signs of phrasal downstep or tonal coupling, were nevertheless clearly smaller than the speech unit as a whole.

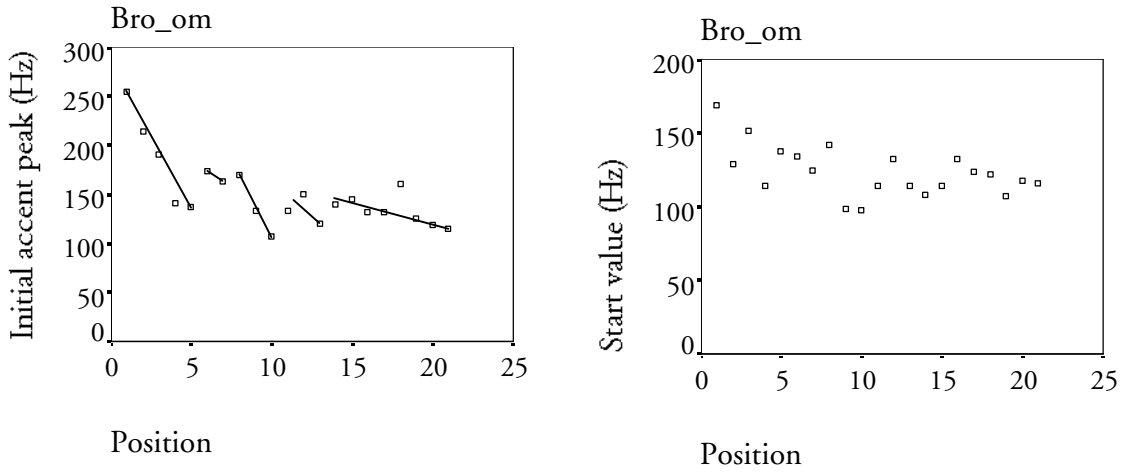


**Figure 5.3** *Initial accent peak (to the left) and start value (to the right) (in Hz) as a function of the position in the speech unit (Bro_om). Five possible downstepped sequences of prosodic phrases are marked based on the values of the phrase-initial accent peaks. Several other analyses are possible.*

In Figure 5.3, we can observe three or four resets of F0 (where the phrase initial accent peaks are clearly higher than the preceding phrase's initial accent peak), and consequently four or five declination ramps or downstepped/tonally coupled sequences of prosodic phrases.

Figure 5.4 gives an example of the F0 contours of three prosodic phrases that can be regarded as a sequence of downstepped prosodic phrases based on their phrase-initial accent peaks' F0 value (phrases 8, 9 and 10 in Figure 5.3 above). It is clear that there is a downward scaling of each prosodic phrase's register relative to the preceding phrase.

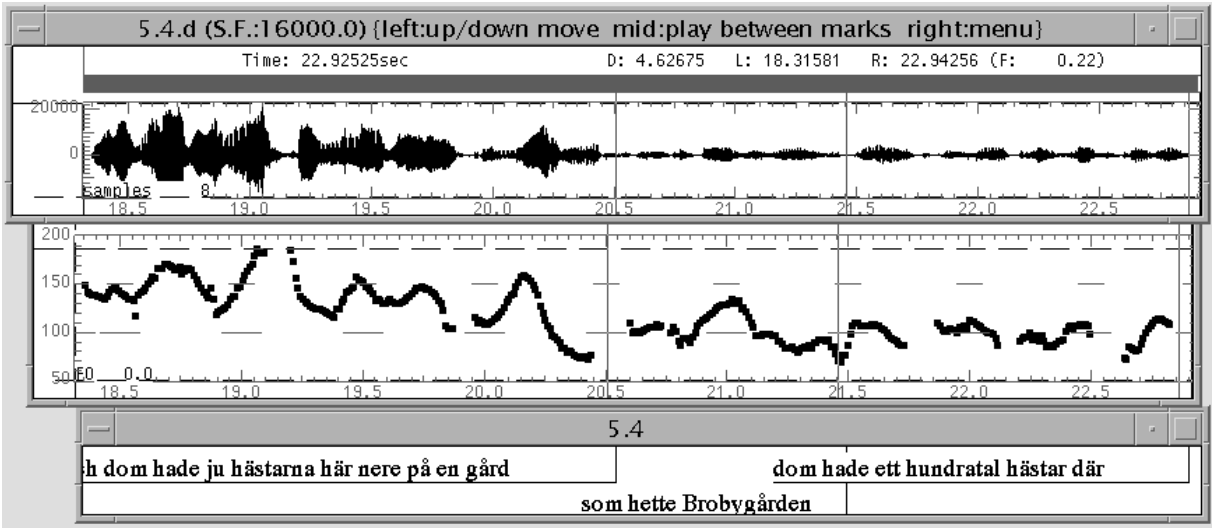


**Figure 5.4** *Wave form and F0 contour of the three prosodic phrases* dom hade ju hästarna här nere på en gård *'they had the horses on a farm down here',* som hette Brobygården *'that's called Brobygården' and* dom hade ett hundratals hästar där *'they had about a hundred horses there' (Bro_om).*

A more detailed analysis of each speech unit separately in which the F0 starting points of adjacent prosodic phrases can be compared is clearly needed. Such an analysis can be made by studying scatter plots like the one above (in Figure 5.3 above), where the F0 starting points are plotted as a function of their position in the speech unit.

Let us examine another speech unit: the speech of the younger speaker from Norra Rörum (*Nro_ym*). This speech unit contains 17 prosodic phrases, although, due to vocal fry, we have no reliable records of the seventh and seventeenth phrase's F0 starting points. As shown in Figure 5.5, despite the lack of a significant decrease in F0 across the entire unit, we nevertheless can see traces of some sort of tonal coherence signaling among the phrases. Three sequences of downstepped phrases can be observed; the six (or seven) first prosodic phrases pertain to one such sequence, the next four (or five) prosodic phrases to another, and the last four to a third sequence of downstepped phrases. It is difficult to say to which sequence of phrases the sixteenth phrase should be grouped since we have no reliable record of the following prosodic phrase's F0 starting point, but given its high peak, it is likely

that phrases number sixteen and seventeen constitute a downstepped sequence on their own (or at least the beginning of such a sequence).

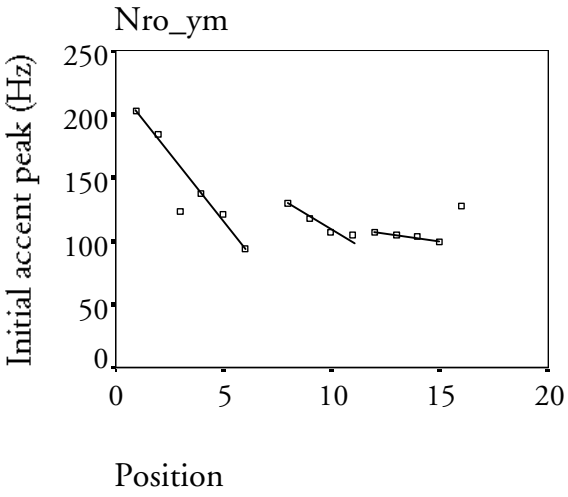


**Figure 5.5** *Initial accent peak (in Hz) as a function of the position in the speech unit (Nro_ym). Three possible downstepped sequences of prosodic phrases are marked. Other analyses are possible.*

Listening to the speech, and comparing our grouping of the prosodic phrases above (which is based only on the F0 values of the phrase-initial accent peaks) with the boundary strengths perceived by the two expert transcribers, we can get an insight into the perceptual relevance of the observed decrease in F0 between successive phrases.

It is not difficult to see the similarities between the grouping of phrases made above based on the initial accent peaks' F0 values and the transcribers' transcriptions. The seventh prosodic phrase (which we were not able to group due to the lack of measurements) is perceived to cohere with the first six, and after it, a strong boundary is perceived by both transcribers. As regards the boundaries after the next two stipulated sequences of phrases (4 + 4 phrases), the transcribers show some disagreement; both perceive the boundaries in question but one transcribed them as strong (see the transcription in (5a)) and the other transcribed them as weak.

Some similarities can also be found between the grouping of the prosodic phrases and the textual structure of the speech, see (5a). In the transcription in (5a) prosodic phrases between which strong boundaries were perceived are separated by an empty line. The eighth prosodic phrase introduces a new topic, whereas the twelfth and sixteenth phrase are continuations of what has previously been said. In the twelfth prosodic phrase, the expressions *vi* 'we' and *varandra* 'each other' refer back to persons already introduced into the discourse, thereby linking the phrase

with the preceding phrases. The sixteenth prosodic phrase is the main clause to the preceding temporal clause.

(5a)
$_{1}$sen idag när jag idag | $_{2}$jobbar jag i skogen | $_{3}$jag har jag har ju skog liksom en del själv också | $_{4}$så jag sysslar en rätt stor del med det | $_{5}$och sen har jag en en scotare | $_{6}$så jag kör entreprenad | $_{7}$lite grann med den också |

$_{8}$sen direkt kan(ske) på sommarmånaderna | $_{9}$då det kanske inte är så mycket med det | $_{10}$då hjälper jag även min kusin | $_{11}$den kusinen som jag liksom gick lite grann i gick i skogen ihop och |

$_{12}$ja vi har fortsatt att hjälpa varandra och | $_{13}$han har maskinstation inom lantbrukssidan | $_{14}$så då blir det lite grann med den biten | $_{15}$på sommaren ibland när han behöver lite extra hjälp så |

$_{16}$hoppar jag in lite på det också | $_{17}$kan man säga | (*Nro_ym*)

'$_{1}$today when I today | $_{2}$I worked in the woods | $_{3}$I have I have like some woods myself too | $_{4}$so I work quite a bit with that | $_{5}$and then I have a scooter | $_{6}$so I do contract driving | $_{7}$a little with it too |

$_{8}$then in the summertime perhaps | $_{9}$when there's maybe not so much happening there | $_{10}$then I also help my cousin | $_{11}$the cousin that I like went together a little went together with in the woods and |

$_{12}$yeh we still help each other and | $_{13}$he runs a garage for farm machinery | $_{14}$so there's a little to do there | $_{15}$in the summer sometimes when he needs a little extra help then |

$_{16}$I hop in there a little too | $_{17}$so to speak |' (*Nro_ym*)

As will be discussed in the next section, the relationship between the perceived strength of a boundary and the degree of tonal coherence existing between the prosodic phrases on either sides of the boundary (as measured by the size and direction of the difference in starting points) is not always as straightforward as the analysis above would suggest.

Based on the F0 values of the phrase-initial accent peaks, we suggested a 5 + 2 + 3 + 3 + 8 grouping of the phrases in the speech unit depicted in Figure 5.3 above (*Bro_om*). Indeed, after the first five phrases, both transcribers perceive a strong boundary. Strong boundaries are also perceived after phrases 7 and 21. However, no strong boundary is perceived after the tenth or thirteenth phrase. Strong (strongly or extra strongly marked) boundaries are, on the other hand, perceived by

both transcribers after the second, eleventh and seventeenth prosodic phrase, see (5b). Prosodic phrases between which both expert transcribers have perceived strong boundaries are separated by an empty line. We would not have been able to predict the expert transcribers' grouping of the phrases based on our observations of the phrase initial accent peaks' relations to each other. Clearly there are other cues than F0 lowering or resetting that affect the transcribers' perception of boundary strength.

(5b)

$_{1}$men jag var med | $_{2}$och red ut hästarna |

$_{3}$för att de skulle till exempel upp på Glimåkra | $_{4}$och de skulle upp på Tydinge | $_{5}$och ta vissa scener |

$_{6}$så red jag alltid han Bror Byglers häst | $_{7}$för han bodde här intill |

$_{8}$så att och dom hade ju hästarna här nere på en gård | $_{9}$som heter Brobygården | $_{10}$dom hade ett hundratal hästar där | $_{11}$som användes av dom |

$_{12}$(det) var ju militärer där nere som | $_{13}$höll ordning på dom ju | $_{14}$så att det var bara att sitta upp och så | $_{15}$åka ut på | $_{16}$skådespelarna dom red ju inte | $_{17}$utan just för det tillfället satt dom på hästen ju |

$_{18}$sen var det ju andra | $_{19}$när det var några riktiga scener och sånt | $_{20}$så var det ju andra som | $_{21}$red ju | (*Bro_om*)

'$_{1}$but I was there | $_{2}$and rode the horses out |

$_{3}$since they were to be taken for example up to Glimåkra | $_{4}$and they were to be taken up to Tydinge | $_{5}$and shoot certain scenes |

$_{6}$I always rode his – Bror Bygler's – horse | $_{7}$since he lived right nearby |

$_{8}$so that and they had the horses down here on a farm | $_{9}$that's called Brobygården | $_{10}$they had about a hundred horses there | $_{11}$that were used by them |

$_{12}$there were soldiers down there that | $_{13}$kept them in order | $_{14}$so that you just had to get on and | $_{15}$ride out | $_{16}$the actors they didn't ride | $_{17}$but just for a particular occasion they got on a horse |

$_{18}$then there were others | $_{19}$when there were real scenes and such | $_{20}$then there were others who | $_{21}$rode |' (*Bro_om*)

One such cue is pause length. Phrase 11 in (5b) above has been grouped with phrases 8, 9 and 10 (see Figure 5.4 above) by the expert transcribers (in the sense that all boundaries between the above-mentioned phrases are perceived as weak and

the boundary after phrase 11 as strong). The F0 value of phrase 11's initial accent peak does not in itself give an unambiguous indication to such an interpretation, since it is higher than the preceding phrase-initial peak, but lower than the following. The fact that the transcribers agree on this grouping (by classifying the boundary preceding it as weak and the boundary following it as strong) indicates that there nevertheless are cues in the speech signal that allow listeners to make such an interpretation. It can be hypothesized that the relatively shorter pause (1.3 seconds) preceding phrase 11 than that which follows (3.4 seconds) phrase 11 constitutes such a cue, see Figure 5.6. There is of course nothing to prevent the listeners from also using clues provided by syntax.

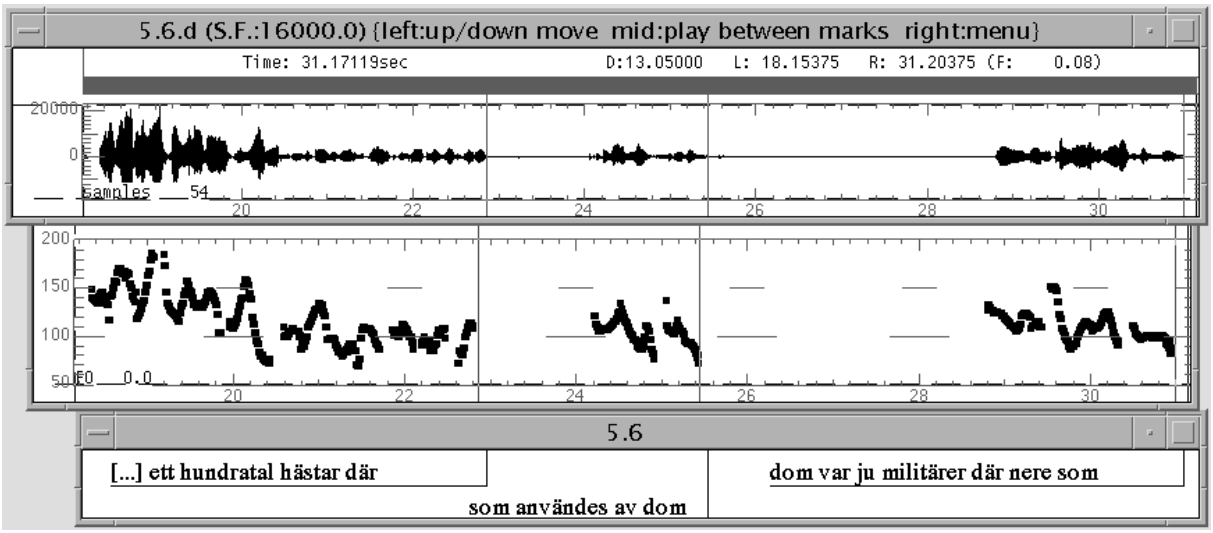


**Figure 5.6** *Wave form and F0 contour of the five prosodic phrases* dom hade ju hästarna här nere på en gård *'they had the horses on a farm down here'*, som hette Brobygården *'that's called Brobygården',* dom hade ett hundratals hästar där *'they had about a hundred horses there',* som användes av dom *'that were used by them' and* dom var ju militärer där nere som *'there were soldiers down there that' (Bro_om).*

Syntactic boundary type has a clear influence on how strong boundaries are produced and/or perceived in the material. A common pattern in e.g. the multi-clause sentences involves a weak boundary at internal clause boundaries and a strong boundary at sentence boundaries. Some examples of this pattern are given in (5c). In all examples, both expert transcribers perceived strong boundaries at the sentence boundaries marked with a '||'.

(5c)

$_{1}$och så hade han ju hade han en hjälpare | som som kom hit också med honom ju || (*Bar_om*)

$_{2}$för då jobbar man fyra dagar | och är ledig tre liksom | så man får alltid en extra dag och sen || (*Bar_ym*)

$_{3}$så red jag alltid han Bror Byglers häst | för han bodde här intill || (*Bro_om*)

$_{4}$och sen får man ju försöka lägga dom | så att dom orkar upp dagen efter va || (*Bro_ym*)

$_{5}$då hade vi kvar göddjuren | och ökade på svinproduktionen istället || (*Lod_om*)

$_{6}$varenda gång man gick till frukost eller middag | så var man lika förvånad över vad är det för väder ute || (*Lod_ym*)

$_{7}$min morfar skulle säga moster till henne | men hon var lika gammal som han || (*Nro_om*)

$_{8}$och sen har jag en scotare | så jag kör entreprenad | lite grann med den också || (*Nro_ym*)

$_{9}$den ene var i Gävle och studerade | och den ene var i Uppsala || (*Oss_om*)

$_{10}$drivande hund det är egentligen det är ju tax och drever och sådant här | och det håller i regel på i flera timmar || (*Oss_ym*)

‘$_{1}$and then he had a helping hand | who also came here with him ||’ (*Bar_om*)

‘$_{2}$because then you work four days | and are off like three | so you always get an extra day and then ||’ (*Bar_ym*)

‘$_{3}$I always rode his – Bror Bygler’s – horse | since he lived right nearby ||’ (*Bro_om*)

‘$_{4}$and then you have to try to get them to bed | so that they manage to get up the next day ||’ (*Bro_ym*)

‘$_{5}$then we kept the beef cattle | and increased the pig production instead ||’ (*Lod_om*)

‘$_{6}$every time you went to breakfast or dinner | you were always just as surprised over the weather out there ||’ (*Lod_ym*)

‘$_{7}$my grandfather had to call her auntie | but she was just as old as he ||’ (*Nro_om*)

‘$_{8}$and then I have a scooter | so I do contract driving | a little with it too ||’ (*Nro_ym*)

‘$_{9}$one of them was in Gävle studying | and the other one was in Uppsala ||’ (*Oss_om*)

‘$_{10}$driving dogs that’s actually that’s dachshunds and drevers and dogs like that | and they can as a rule go on for several hours ||’ (*Oss_ym*)

Since the classification of boundaries as either clause or sentence boundaries is far from unproblematic, we will make no attempts to estimate how frequent the above-mentioned pattern is in our material. Coordinated main clauses and sentences introduced by conjunctions functioning as discourse markers are very frequent. Since the interpretation of words like *och* 'and' or *men* 'but' as coordinators of main clauses or as discourse markers is partly dependent on prosody (Horne, Hansson, Bruce, Frid and Filipsson 2001), e.g. the strength of the preceding boundary, we risk introducing circularity in our analysis (if prosody is used in the classification of boundary type, then boundary type will have prosodic correlates).

We will therefore only conclude that the relationship between syntactic boundary type and boundary strength is not a one-to-one relationship. Although the large majority of the boundaries perceived as strong by both expert transcribers occur at sentence boundaries, examples of sentence boundaries that are only weakly marked abound (see above), and there are also examples of strongly marked internal clause boundaries in the material.

### 5.3.2 Signals of coherence vs boundary signals

A comparison between the two expert transcribers' labeling of boundary strength in the male subjects' speech indicate that agreeing on boundary strength is more difficult than agreeing on whether or not a given word is followed by a boundary or not. A closer review of the expert transcribers' labeling also reveals that there is no simple relationship between the perceived strength of a boundary and the degree of tonal coherence existing between the prosodic phrases on either side of the boundary (as measured by the extent and direction of the difference in starting points). It is not the case that tonally coherent prosodic phrases are always perceived to be separated by weak (prosodic phrase) boundaries. Nor is it the case that phrases between which the starting point is reset are always perceived as divided by strong (prosodic utterance) boundaries. This is most likely due to the presence or lack of other strong boundary signals that modify or function to complement the degree of perceived boundary strength. Pauses are likely to be one such strong boundary signal. The presence of a pause can be expected to modify and decrease the perceived degree of coherence between two prosodic phrases, whereas the lack of a pause between two phrases probably increases the amount of coherence perceived between the phrases. Note that we are assuming that the degree of perceived boundary strength and the degree of perceived coherence across the boundary are related.

Conflicts between strong boundary signals (such as a long pause) and coherence signaling cues (such as a decrease in the F0 starting point between successive phrases) often arise in spontaneous speech. By 'conflict' we mean that the boundaries in question are not easily classified as to boundary strength within the Swedish base prosody system due to the presence of both coherence and boundary signaling cues[4]. However, the possibility to use conflicting cues causes no problems in the speech situation, but should rather be seen as an asset to the speaker. The speaker can, e.g., make an addition to what (s)he has said although its end has already been marked by a pause. (S)he can do this by making the addition cohere intonationally with the preceding speech, see Figure 5.7. Here the addition *på vintrarna* 'in the winters' follows a pause and the steep F0 fall to a low end point in the focally accented phrase-final word *säd* 'grain'. It is nevertheless clearly understood as an addition to the preceding utterance rather than as the beginning of a new utterance (which is also possible syntactically).

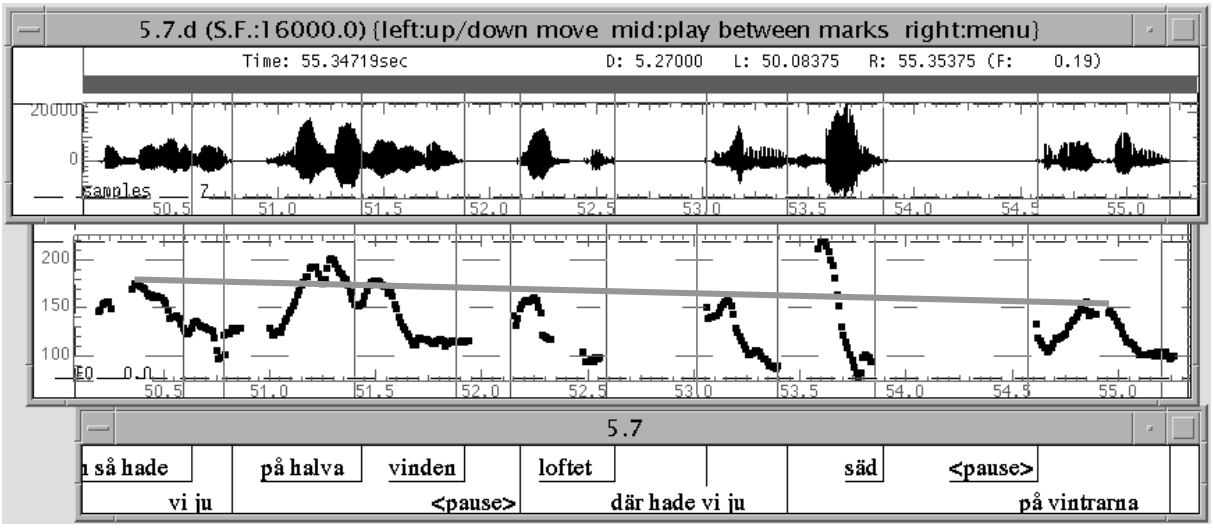


**Figure 5.7** *Speech wave and F0 contour of* och så hade vi ju på halva vinden – loftet – där hade vi ju säd på vintrarna *'and we had in half the attic – the loft – there we had grain in the winters' (Oss_ow). A line has been drawn through the phrase-initial accent peaks to illustrate the decrease in F0 across phrases.*

The fact that the F0 decrease of phrase-initial accent peaks between *på vintrarna* and the preceding phrase is similar to the F0 decrease occurring between the

[4] In the next chapter, we will investigate the relationship between F0 reset, pausing and phrase-final lengthening on the one hand and perceived boundary strength on the other, and discuss the conflicts that arise when classifying boundary strength in spontaneous speech within the Swedish base prosody system.

parenthetical expression[5] *loftet* 'the loft' and its preceding phrase, as well as between the continuation *där hade vi ju* 'there we had' and the parenthetical *loftet*, indicates to us that intonational coherence signaling can be achieved without lookahead. It seems reasonable to assume that the speaker had a continuation planned when inserting the parenthetical expression. Unlike the addition *på vintrarna*, the planned continuation shows several signs of prosodic coherence with the preceding speech (e.g. a relatively short pause). However, as regards the tonal coherence signaling among the phrases, no evident differences can be found between how the preplanned continuation is made to cohere with the preceding speech and how the postplanned addition is made to cohere.

Analyses of additions like the one above in Figure 5.7, are interesting from a planning point of view. How are postplanned additions incorporated into the preceding speech? Is the speaker able to make it cohere intonationally in a way similar to prosodic phrases occurring in more fluent parts of speech? If so, then we must assume that the tonal coherence signaling under investigation requires no lookahead. The example in Figure 5.7 above lends some support to such a conclusion.

In Figure 5.8, the speaker ends her story about the day when she told her employer that she was not the one who used up all the cream by saying that she was never accused of that again (*de de det sa hon aldrig mer till mig* 'she never said tha tha that to me again'). The end is strongly marked prosodically. However, apparently the speaker feels that a clarification is needed, and makes the addition *om detta här* 'about this matter'. Unlike after phrase-internal pauses, the first accent's peak after the pause is not lower than the last accent peak before the pause. A new prosodic phrase is clearly started[6], but it is made coherent with the preceding by the speaker who chooses a F0 starting point that is lower than the starting point of the preceding phrase. Had the speaker planned the prepositional phrase in question beforehand, it could easily been wrapped into the same prosodic phrase as *det sa hon aldrig mer till mig* 'she never said that to me again'.

[5] The parenthetical expression differs from a so-called 'modification repair' (see *2.3.1*) in that it is set aside as a prosodic phrase on its own.

[6] The strong marking of the preceding prosodic phrase's end clearly cannot be entirely overridden by any coherence signaling cues.

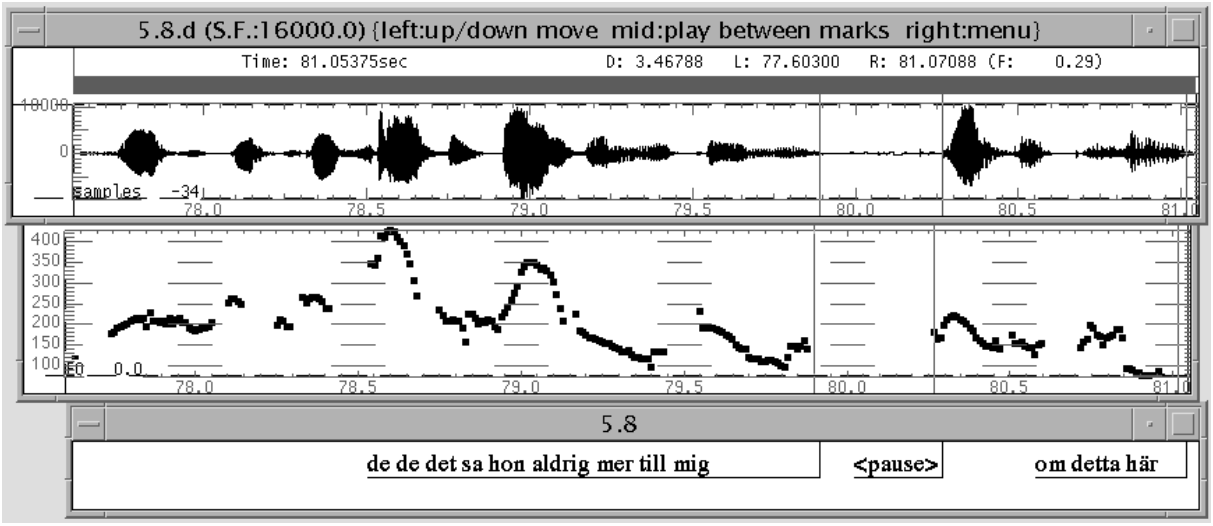


**Figure 5.8** *Wave form and F0 contour of the two prosodic phrases* de de det ˈsa hon ˈàldrig mer ˈtill mig *'she never said tha tha that to me again', and* ˈom detta ˈhär *'about this matter' (Bro_ow).*

There are several examples of adjunct elements that are set aside as separate prosodic phrases in the material with pauses preceding them. Like the pauses occurring at sentence boundaries (Goldman-Eisler 1972), the pauses preceding these additions likely reflect the externalisation of a thought. The strong prosodic marking of the end of the phrase preceding the adjunct element including the presence of a pause indicate to us that these sentences were not entirely planned ahead.

In many cases, the additions (e.g. prepositional phrases) are syntactically ambiguous in the sense that they could either belong to the preceding or following sentence. In the cases examined so far, prosody clearly helps the listener to disambiguate these structures i.e. to understand to what sentence the addition belongs. However, the plan-as-you-go basis of spontaneous speech and the apparent possibility speakers have to produce tonally coherent sequences of phrases without lookahead, open many possibilities for the speaker. The speaker may e.g. make changes to the original plan as he or she goes along, thereby producing syntactically "impossible" – but communicatively perfectly acceptable – structures. Let us examine a few examples in detail.

In (5d), the speaker talks about the development in farming and how he kept his farm until 1978 when he handed it over to his son. The year 1978 is pronounced as a phrase on its own after a pause where the speaker seems to think for a while about exactly what year it was. When pronouncing the year, the speaker would appear to already have a continuation planned and the year 1978 gets the double function of being both part of the sentence-final prepositional phrase *till 1978* 'to 1978' and

the sentence-initial temporal adverb in the following sentence (*1978 så överlämnade jag det till min son* 'in 1978 I turned it over to my son'). The prosodic phrase *1978* is thus the central element in what Linell (in preparation) terms an 'apo-koinou' construction. The F0 tracings of the three prosodic phrases are shown in Figure 5.9.

(5d)
$_1$och sedan fortgick det väl så här fram till | $_2$nittonhundrasjuttioåtta | $_3$så överlämnade jag det till min son och | (*Lod_om*)

'$_1$and then things continued on like this until | $_2$(in) nineteen seventy-eight | $_3$I turned it over to my son and |' (*Lod_om*)

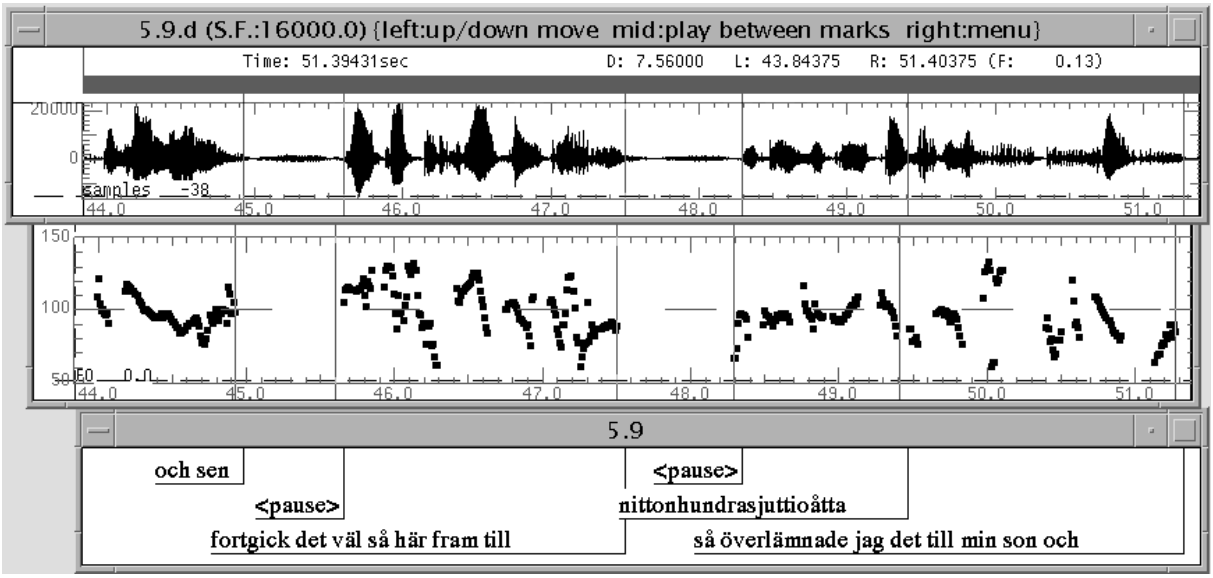


**Figure 5.9** *Wave form and F0 contour of the three prosodic phrases* och sen fortgick det väl så här till *'and then things continued on like this until',* nittonhundrasjuttioåtta *'(in) nineteen seventy-eight', and* så överlämnade jag det till min son och *'I turned it over to my son and' (Lod_om).*

Despite the syntactic evidence of replanning during the execution of the utterance, the tonal coherence is not different from that which can be observed in fluent and syntactically well-formed parts of the speaker's speech.

In (5e), the speaker talks about the situation of maids in Sweden in her youth: they had to eat in the kitchen while the family got to eat in the dining-room. While the first prosodic phrase is separated from the following by a weak boundary, the end of the second phrase is more strongly marked prosodically. The two prosodic phrases are semantically coherent and syntactically coordinated and the presence of a weak boundary between the clauses and a stronger prosodic marking after them (the

sentence boundary) is therefore both semantically and syntactically motivated, see Figure 5.10. However, the speaker chooses to make an addition (*när dom hade ätit* 'when they had eaten'). She makes clear that the maids had to wait to eat until after the family had eaten. The end of this prosodic phrase is also strongly marked prosodically (the new end of the sentence). The phrase is furthermore tonally coherent with the preceding. The strength of the preceding phrase boundary is thus presumably reevaluated by the listener, and the addition is perceived as a continuation, see Figure 5.11. The speaker then chooses to clarify the situation further, by adding the prosodic phrase *så fick hon äta resterna* 'then she got to eat the leftovers'. The temporal clause 'when they had eaten' can now be interpreted as either being a part of the preceding or the following sentence. Because the last clause in (5e) cannot stand on its own syntactically, only one syntactic analysis is possible: the sentence boundary precedes the temporal clause. Prosodically, however, this analysis is not motivated (see Figure 5.12): each prosodic phrase is a continuation of the preceding, and the temporal clause functions first to modify the first, and then the second sentence (or sentence-like unit).

(5e)
$_1$där åt ju familjen inne i matsalen | $_2$och sen jungfrun hon fick nöja sig med att äta i köket | $_3$när dom hade ätit | $_4$så fick hon äta resterna | (*Bro_ow*)

'$_1$the family ate in the dining-room | $_2$and then the maid had to be satisfied with eating in the kitchen| $_3$when they had eaten | $_4$then she got to eat the leftovers |' (*Bro_ow*)

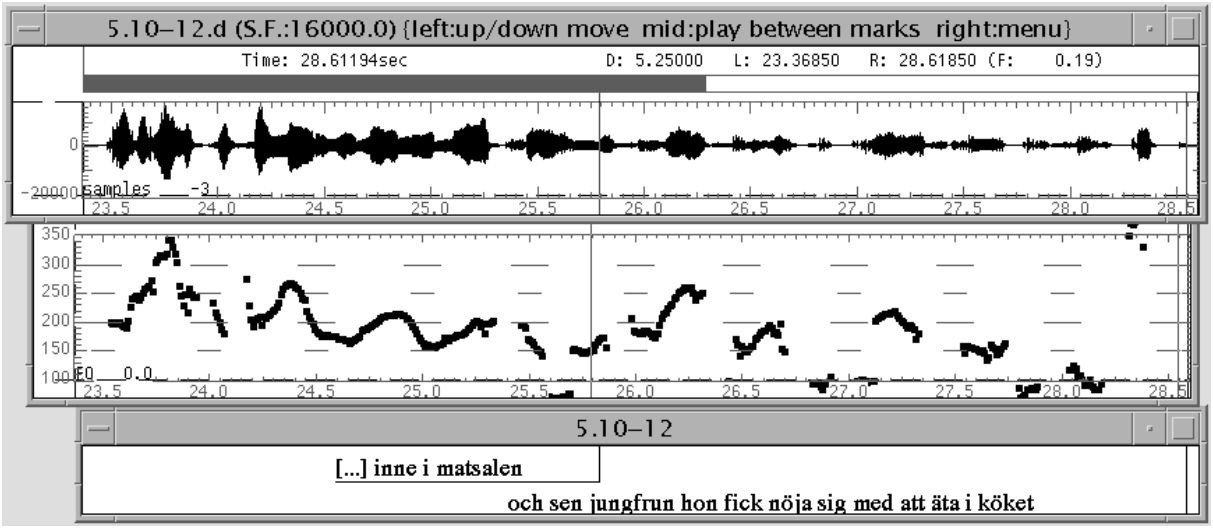


**Figure 5.10** *Wave form and F0 contour of the two prosodic phrases* där åt ju familjen inne i matsalen *'the family ate in the dining-room', and* och sen jungfrun hon fick nöja sig med att äta i köket *'and then the maid had to be satisfied with eating in the kitchen' (Bro_ow).*

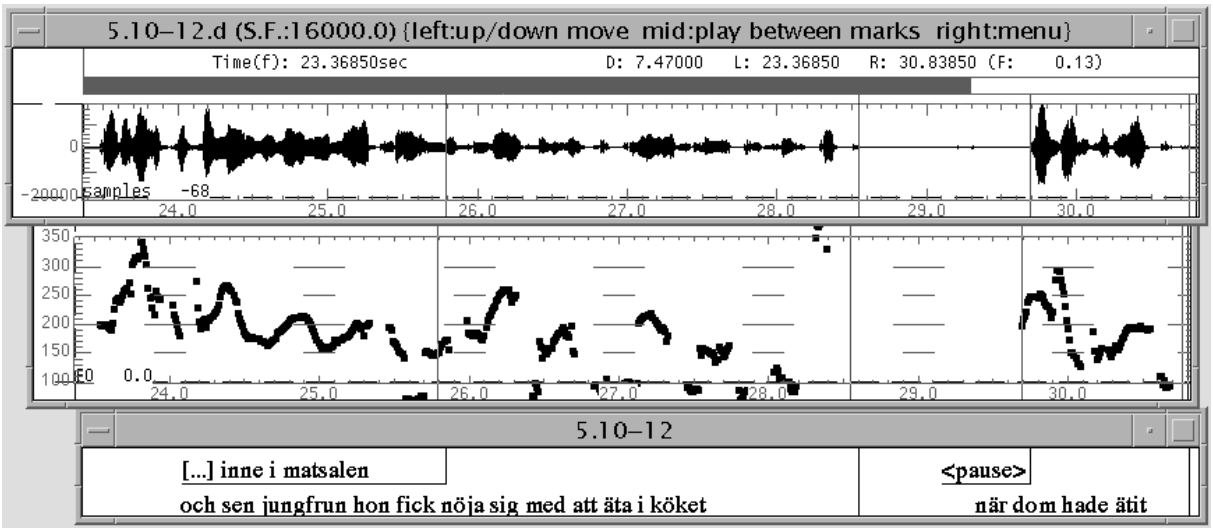


**Figure 5.11** *Wave form and F0 contour of the three prosodic phrases* där åt ju familjen inne i matsalen *'the family ate in the dining-room',* och sen jungfrun hon fick nöja sig med att äta i köket *'and then the maid had to be satisfied with eating in the kitchen' and* när dom hade ätit *'when they had eaten' (Bro_ow).*

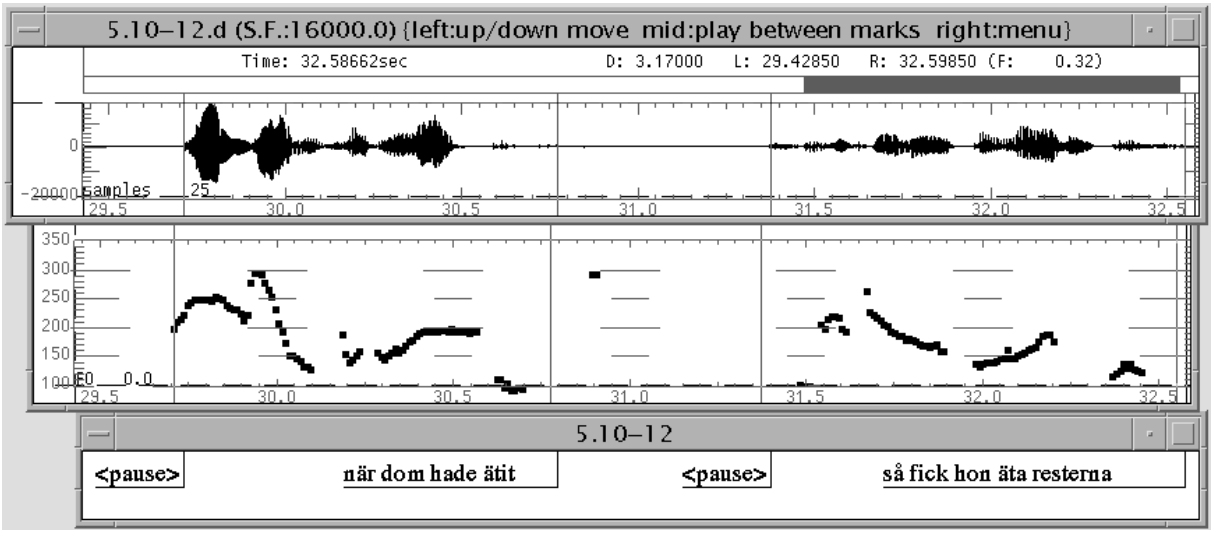


**Figure 5.12** *Wave form and F0 contour of the two prosodic phrases* när dom hade ätit *'when they had eaten' and* så fick hon äta resterna *'then she got to eat the leftovers' (Bro_ow).*

We believe that this flexibility of spontaneous speech, observed in both (5d) and (5e), reflect a plan-as-you-go basis. The tonal coherence that nevertheless can be observed and perceived among the phrases indicates that no lookahead is required to produce downstepped/tonally coupled phrases.

### 5.3.3 Implications for preplanning

What implications does the finding of tonal coherence signaling among prosodic phrases have on our conception of the amount of preplanning involved in prosodic phrasing in spontaneous speech? Liberman and Pierrehumbert (1984) make a distinction between what they term 'hard' and 'soft' preplanning. Hard preplanning refers to such processing that needs to be accomplished prior to the execution of the phrase, while soft preplanning refers to the sort of planning that the speaker may choose to make prior to the execution of the phrase, but can omit.

The adjustment of F0 slope to phrase length as understood and modeled in the original Lund model requires hard preplanning. An obligatory adjustment of the F0 starting point to phrase length, as suggested in the early work on downstep by Bruce (1982a), would also require some lookahead and hard preplanning. However, no evidence to suggest that our speakers do such hard preplanning could be found in the data. That does not mean that the speakers are prevented from planning ahead. What is important here, however, is the finding that they do not always plan ahead, or even frequently make these kinds of adjustments that require lookahead. When adjustments are made, they must be seen as the result of optional – soft – preplanning.

Thus, although the lack of evidence of lookahead is not evidence against hard preplanning, we nevertheless believe that preplanning in spontaneous speech involves no hard preplanning. Some support against hard preplanning can be found in our analyses of postplanned additions. We believe that they demonstrate that well-formed tonal coherence signaling is produced by speakers even in utterances where the speaker plans his or her speech on a plan-as-you-go basis. Of course this conclusion rests on the assumption that we have correctly identified utterances that at their start were not entirely planned in our analysis. Here, we have taken examples of prosodic phrases that are added to an otherwise clearly prosodically terminated topic to be postplanned additions, in many cases an adjunct element such as a prepositional phrase. In some cases, syntax has also served to indicate that an utterance was not entirely planned ahead or that it had been reorganized during production.

That the speaker would have "the choice of continuing a downdrift" in spontaneous speech was hypothesized already in Bruce (1982b: 285).

## 5.4 Summary

Based on our observations of groups of prosodic phrases with uninterrupted decrease in F0 (observed here in the F0 starting points), we believe that intonation is also used to signal coherence within larger domains than the prosodic phrase in spontaneous Swedish.

Whether the downward scaling of phrase-initial accent peaks and register in successive prosodic phrases should be regarded as the result of phrasal downstep, as suggested by van den Berg *et al.* (1992), tonal coupling, as suggested by Bruce (1982b) or some other mechanism, we cannot say. Our main concern in the study reported on here was simply to determine if decreasing F0 over the course of several prosodic phrases is used to signal coherence in spontaneous speech.

By examining a number of groups of downstepped/tonally coupled prosodic phrases in more detail, we concluded that coherence is perceived among the tonally coherent phrases, and consequently that the lowering of phrase-initial accent peaks likely is perceptually relevant. Naturally, there may be other cues that coexist with the decrease in F0 that help, and perhaps are more important in the signaling of coherence among prosodic phrases. The present analysis does not allow us to draw any conclusions as to F0's unique contribution to the perceived degree of coherence. The fact that neither does lowering F0 starting points between prosodic phrases invariably give rise to the perception of coherence across the phrase boundary (i.e. a weak prosodic boundary) nor does resetting invariably give rise to the perception of a strong boundary between the phrases, suggests the presence of other cues that also affect the degree of perceived coherence.

We also showed that the observed decrease in F0 starting points tends to occur within semantically and syntactically coherent units, e.g. between clauses included in the same sentence.

As far as our previous conclusions about the course of F0 in the prosodic phrase are concerned (see chapter four), a revision is needed. We believe that the F0 starting points are adjusted to signal the relation between the upcoming prosodic phrase and the preceding. An upward scaling of the upcoming phrase's register signals discontinuation, and a downward scaling signals continuation.

The initial accent peak's value is thus chosen based most likely on a combination of expressive and discourse factors, and thereafter the phonetic values of the accents' peaks and valleys in the prosodic phrase are chosen taking only the previous

accent's values into consideration. Lookahead can be employed but is not required, either in the production of the accentual structure of the prosodic phrase, or in the production of the phrasal structure of discourse.

CHAPTER 6

# Boundary strength

## 6.1 Introduction

After having considered aspects of both boundary signaling and coherence signaling in prosodic grouping, we now turn to the signaling of boundary strength in the phrasing of spontaneous speech. In the present chapter, we follow up on the finding that intonation alone cannot explain the perceptual impression of coherence among prosodic phrases (see chapter five) by presenting the results of two perception experiments.

The perception experiments described in the present chapter have two purposes: 1) to relate perceived boundary strength to three known cues for prosodic phrasing in Swedish (pausing, F0 reset and final lengthening), and 2) to investigate whether the established division into two phrasal categories or constituents (the 'prosodic phrase' and the 'prosodic utterance') has empirical support in spontaneous Swedish.

### 6.1.1 Perception of prosodic phrasing in Swedish

We already have some knowledge about what cues are relevant for the perception of prosodic phrasing in Swedish. The importance of segment duration and F0, for

example, has been investigated in a series of perception experiments that tested the perceptual relevance of cues observed in read production data (Bruce *et al.* 1993). It was hypothesized and subsequently shown that a shallow F0 valley as part of a downstepping pattern has a connective function signaling coherence within a prosodic phrase, while a deep F0 valley as a break in the downstepping trend has a demarcative function signaling a phrase boundary. It was furthermore shown that increased duration in segments is perceived as a cue for a boundary whereas reduced duration signals coherence. In the experiment, a syntactically ambiguous test sentence with two possible sentence internal clause boundaries was used and the listeners' task was to choose optimal, i.e. disambiguating combinations of the two manipulated parameters for both interpretations. The results revealed that both the boundary and coherence signaling cues were important, and that most listeners relied on a combination of duration and F0 for their decisions.

The relative importance of the different features of the boundary signaling 'deep F0 valley' for the percept of phrasing was investigated in House (1990). In a series of perception experiments, House showed that a combination of perceived pitch fall on the final phrase element and a perceived pitch level difference at the phrase boundary functions as a powerful cue for phrasing. Although tonal movement configurations were shown to be important for the percept of phrasing on their own, it was the difference between tonal levels (average F0) of successive syllables which provided the strongest single phrase boundary cue, and the combination of both cues that functioned as the most powerful signal for phrasing.

By 'powerful' and 'strong' is meant that a large proportion of the listeners' responses reflected that the cue in question had been perceived as signaling the presence of a phrase boundary. Whether or not the number of responses can be claimed to relate to the strength of the boundary is not clear. It seems reasonable to assume that a boundary that is perceived by a large number of listeners in some sense is stronger than a boundary that is perceived by few listeners, and if this is so, then we may speculate that both type and number of cues used have an effect on a boundary's perceived strength. The extent of the various cues used, e.g. the amount of final lengthening or the extent of the F0 reset, is also likely to have effects on the perceived strength of the boundary. The experiments undertaken by House (1990) were nevertheless not designed to test different cues' effect upon perceived boundary strength, and the claim of increased degree of perceived boundary strength with an increased number of boundary cues and with an increased amount of the various cues still requires empirical testing.

Some evidence to suggest a relationship between perceived boundary strength and the amount of different phrasing cues used comes from Horne *et al.* (1995). They investigated the relationship between different prosodic category boundaries – prosodic words (PWd), prosodic phrases (PPh) and prosodic utterances (PU) – and final lengthening and pause duration in acoustic registrations of production data (radio broadcasts). Within the hierarchical model of prosodic constituents generally assumed for Swedish, the three categories PWd, PPh and PU are associated with three different degrees of boundary strength. Horne *et al.* (1995) examined how these three degrees of boundary strength and a fourth 0-boundary are related to the presence and amount of final lengthening and pause/silent interval duration[1]. The results showed that pause duration increases as the rank of the boundary becomes higher, i.e. with increasing boundary strength, but that pauses are most intimately tied to the higher-level constituent boundaries (phrase and utterance boundaries). The domain of final lengthening was also concluded to be the prosodic phrase and prosodic utterance, although no conclusive evidence to support increasing final lengthening with increasing boundary strength was found. Segment duration (amount of final lengthening) even proved to be negatively correlated with pause duration at the lower-level constituent boundaries.

In another perception experiment (House, Hermes and Beaugendre 1998), the interrelationship between cues for accentuation and phrasing in Swedish was investigated by systematically shifting the timing of rising and falling pitch movements in the stimuli. It was shown that the same tonal cues can function both as cues for accentuation and phrasing. After having determined the location of the 'accentuation boundary' (House, Hermes and Beaugendre 1997) – the point for onsets of pitch[2] movements where the percept of accentuation shifts from one syllable ($S_n$) to the next ($S_{n+1}$) – the location of the 'phrase boundary' was established. The 'phrase boundary' marks the point where the pitch jump (tonal level difference) or onset of movement shifts from cuing the beginning of the new phrase on $S_n$ to cuing the beginning of the new phrase on $S_{n+1}$. A comparison of the locations of the 'accentuation boundary' and the 'phrase boundary' revealed that the 'phrase boundary' location was earlier in the syllable $S_n$ than the 'accentuation boundary'. This finding was interpreted as suggesting that onset of movement features are stronger cues for accentuation than for phrasing. The perceived onset of a rise at the end of the vowel of $S_n$ would cue an accent on $S_n$ but be perceived as a

[1] English listeners have been shown to expect longer preboundary syllable durations as the rank of the phonological boundary becomes higher (Gussenhoven and Rietveld 1992).

[2] Here 'pitch' is used synonymously with F0.

jump cue for phrasing and cue the beginning of a new phrase on $S_{n+1}$. Jump cues were consequently argued to be stronger cues for phrasing than accentuation. These results are consistent with findings in House (1990) where it was shown that tonal levels are superior to tonal movement configurations as cues to the perception of phrasing.

In summary, to our knowledge, no attempts have been made to investigate the perception of prosodic boundary strength in Swedish. We nevertheless know, from studies of both production and perception, that a whole constellation of cues is involved in prosodic phrasing in Swedish: F0, pausing and phrase final lengthening. Changes in the acoustic correlates of voice quality and intensity are also known cues to phrasing that due mainly to methodological difficulties have not been examined to the same extent in Swedish (but see Huber 1988). All these cues – changes in F0, pausing, phrase-final lengthening, laryngealization[3] and changes in intensity – may have importance for the perception of boundary strength.

### 6.1.2 Perceived boundary strength and the prosodic hierarchy

As discussed by Ladd (1996), differences in prosodic boundary strength are only possible under the 'Strict Layer Hypothesis' (Selkirk 1984) if they reflect differences of boundary type. Different kinds of boundaries separate different kinds of units and, according to the Strict Layer Hypothesis, any unit at a given level of the hierarchy consists exclusively of units at the next lower level of the hierarchy. In the tonal transcription system for Swedish (Bruce *et al.* 1994), a distinction is made between two boundary strengths and two phrasal categories or constituents above the prosodic word: prosodic phrases (which are delimited by weak, single bar boundaries) and prosodic utterances (which are delimited by strong, double bar boundaries).[4] A prosodic utterance consists of one or several prosodic phrases. Although convincing evidence has been put forward to support the distinction made between three degrees of prominence in the intonation model for Swedish (see section 1.4), the validity of assuming two phrasal categories has not been

[3] 'Laryngealization', 'glottalization', 'creak', 'creaky voice' and 'glottal fry' are terms that are largely used synonymously in the literature (Huber 1988).

[4] In the base prosody transcription system (Bruce 1994) for Swedish (but not in the tonal transcription system for Swedish (Bruce *et al.* 1994)) three degrees of boundary strength above the prosodic word are usually distinguished. The evaluation of the system (Strangert and Heldner 1995a and b), however, has resulted in a proposal to collapse the second and third boundary strength into one.

demonstrated empirically. In the following, we will first discuss the constituents of the prosodic hierarchy generally used in descriptions of Swedish as well as a constituent generally not considered relevant in descriptions of Swedish (the intermediate phrase), and secondly, exemplify some of the problems encountered when describing the hierarchical structure of spontaneous speech.

There is a complex relationship not always acknowledged between the prosodic means used to express prominence in speech and the means used to group speech into prosodic constituents such as prosodic words and prosodic phrases. In the Swedish model for prominence levels (see section 1.4), a distinction is made between three degrees of prominence: stress, accent (Accent I and II) and focal accent (focus). Each and every one of these categories can also be argued to play an important role in prosodic grouping.

It can e.g. easily be shown that the perceptually relevant difference between a two-word prosodic phrase such as *mellan målen* 'between the meals' and the segmentally identical one-word phrase *mellanmålen* 'the snacks' is related to the distinction between stress and accent (see Figure 1.1 in chapter one). Whereas *mellan* carries an accent in both phrases, *målen* is only accented in the two-word phrase. In other words, in the one-word phrase, the first syllable of *målen* is stressed, but due to the lack of an accent not perceived as a word on its own (Bruce 1977). The accent can therefore be claimed to be the defining feature of the 'prosodic word' in Swedish. Note that we by 'accent' mean the second degree of prominence in the Swedish model, i.e. accent I or II, and not phrase accent.

Not only the presence of accents but also their location plays a role for the perceived structure. A three-word unit like *stor mans dräkt* is perceived as one word (a compound) if it contains one accent (i.e. as 'magnate-suit'), and as three (simplex) words if it contains three accents (i.e. as 'big man's suit'), see Figure 6.1, *a* and *b*. However, if it is pronounced with two accents, it is perceived as two words (a simplex and a compound word), and the perceived location of the phrase-internal word boundary depends on the location of the accents, see Figure 6.1, *c* and *d*. If, in addition to *stor*, *mans* carries an accent, then the phrase is perceived as *stor mansdräkt* 'big suit (for a man)', see *c*. On the other hand, if the second accent is carried by *dräkt*, then the phrase is perceived as *stormans dräkt* 'magnate's suit', see *d* (examples from Zetterlund *et al.* 1978). The accent is the head of the word; it is associated with the word's primary stress.

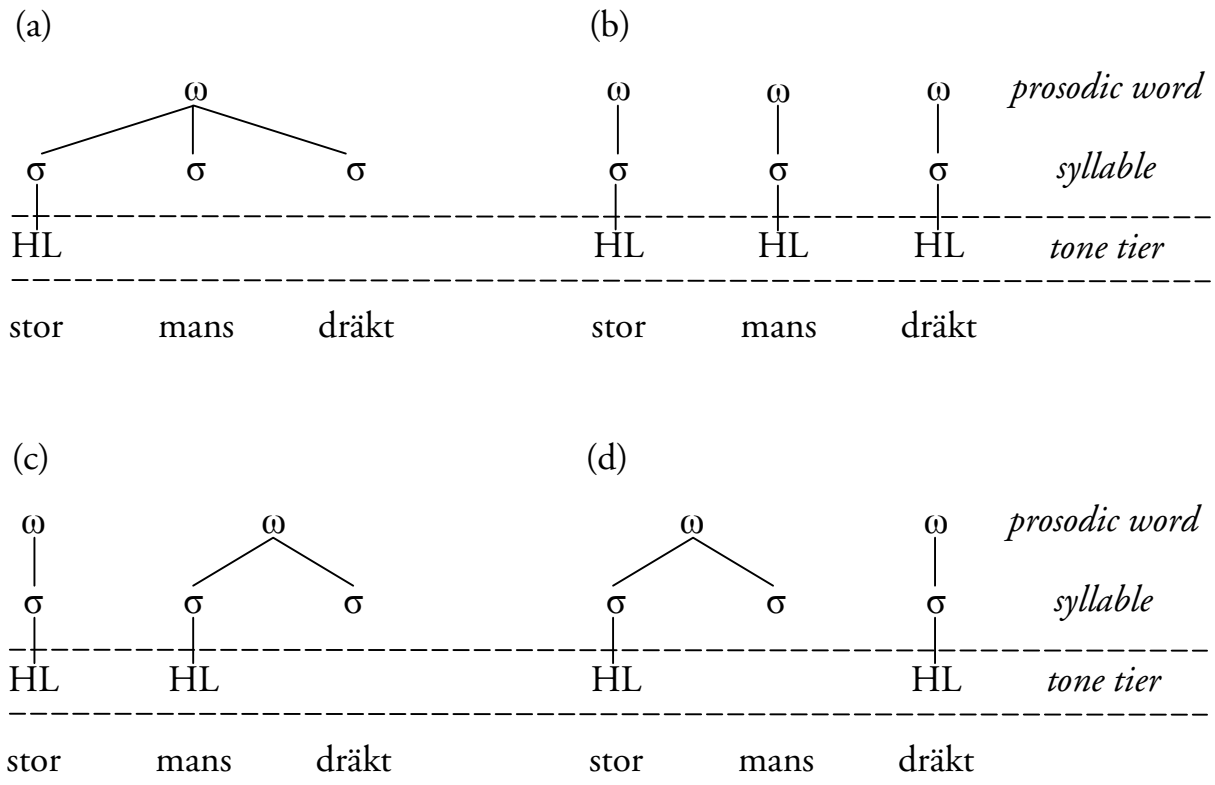


**Figure 6.1** *Representation of associations between syllables and tones. Phrase accents and boundary tones are inserted at the higher phrase level. The words are: a)* stormansdräkt *'magnate-suit', b)* stor mans dräkt *'big man's suit', c)* stor mansdräkt *'big suit' and d)* stormans dräkt *'magnate's suit'.*

The definition of the 'prosodic word' used here is somewhat different from e.g. Nespor and Vogel's (1986) definition of the 'phonological word'[5]. It is, nevertheless, similar to Pierrehumbert and Beckman's (1988: 25) definition of the 'accentual phrase' in Japanese: "the smallest prosodic unit that is well defined in terms of tone pattern; […] the domain of the lexical accent pattern as traditionally described in phonological treatments of Japanese". In contrast to the accentual phrase, the prosodic word, as defined in the present study, has no delimitative peripheral tones that are inserted postlexically. The prosodic word has, however, been proposed to have right-edge boundary tones in Swedish (Horne 1994). The accent unit in Norwegian and the stress group in Danish are other similar units (see section 1.2).

[5] The notions Nespor and Vogel (1986) use to define the phonological word are not the same in all languages. The phonological word is, however, always equal to or smaller than the terminal element of a syntactic tree and, in addition, there is never more than one phonological word in a single stem.

In much the same way as the word accent, the focal (or phrase) accent and its placement, has been suggested to play a role in the grouping of speech into a level of structure above the prosodic word (see e.g. Grice, Ladd and Arvaniti 2000). In Pierrehumbert's (1980) description of English, the phrase accent is the edge tone of the 'intermediate phrase'. However, as will be argued below, no simple relationship exists between focal accent and its position and phrase structure in Swedish.

In a review of the various prosodic hierarchies and constituents that have been proposed in the literature, Wightman *et al.* (1992) note that, the 'intonational phrase' (corresponding loosely to the 'prosodic phrase' in the Swedish intonation model) is a prosodic constituent that is widely accepted among researchers. They define it as a group of words which is delimited in some way as a larger unit of phrasing. Pierrehumbert (1980) maintains that the intonational phrase is a domain for the description of intonation that is not controversial. She notes that it corresponds to what in other works has been called a 'sense group' (Armstrong and Ward 1926, Vanderslice and Ladefoged 1972), 'tone unit' (Crystal 1969), 'tone group' (Ashby 1978, Halliday 1967) and 'breath group' (Lieberman 1967). Although this prosodic group's definition is often somewhat vague and differs slightly between intonation models (see Ladd 1996: 235 for a discussion), one of the most important functions of prosody is to divide the flow of speech into chunks of some sort containing typically a handful of words. Consequently, it is rarely questioned that at least one constituent above the prosodic word is needed in the description of how the stream of speech is structured prosodically. In the following, we term this constituent the 'prosodic phrase' as is customary in Swedish prosody research, thereby stressing its various prosodic correlates (including non-intonational correlates).

Wightman *et al.* (1992) also discuss a second phrasal constituent. It is a constituent upon which fewer researchers agree: the 'intermediate phrase', a level of phrasing between the prosodic word and the intonational phrase. Its hallmark is the presence of a phrase accent that functions as an edge tone. The 'phonological phrase' (Nespor and Vogel 1986) and the 'major phrase' (Selkirk 1984) are similar units that have been proposed in the literature.

The intermediate phrase is generally not assumed to be a relevant phrasal constituent in Swedish. Although in many dialects of Swedish, a phrase accent is often found phrase-finally (prior to the boundary tone where such a tone can be observed), this is not always the case. The phrase accent is found after the word accent fall in words in focal position. Like its English counterpart, the phrase accent

is the tone following the nuclear pitch accent. Focally accented words are, however, not always perceived to be phrase-final, and therefore the presence of a phrase accent is not indicative of a phrase boundary, at least not in the same manner as in English. The phrase accent is not an edge tone in Swedish or, put in another way, it is not peripheral to any phrasal constituent. An exemplification of this fact can be found in the evaluation of the Swedish base prosody transcription system, undertaken by Strangert and Heldner (1995a). In the evaluation, nine expert transcribers (experienced phoneticians and speech researchers specializing in prosody) were asked to mark prominences and boundaries in a short read speech material (233 words). The focal accents on which the largest number of labelers agreed (7 of the 9 labelers) were found on words (*libyska* 'Libyan' and *bombplanen* 'the attack planes') after which none of the labelers perceived phrase boundaries.

Perceptual evaluation of different focus distributions (Bruce *et al.* 1993) has also shown that no reliable, simple relationships exist between focal accent distribution and boundary perception in Swedish. Unlike in English, lexical items in post-focal position are not deaccented in Stockholm Swedish. The phrase accent is therefore not necessarily the last tonal element before the boundary tone. Pierrehumbert and Beckman (1988) relate the fact that accents are allowed to the right of the focal accent to the word accents' status as phonological properties of individual lexical items, and suggest that the Swedish phrase accent be analyzed as the head rather than the edge of the constituent (the intermediate phrase or the intonational phrase). This is not an unproblematic solution, however, since the head of the constituent according to their theory is the stressed syllable of the focused word and this is not the place where the phrase accent is generally found. In compounds in (Stockholm) Swedish, e.g., the phrase accent is aligned with the secondary stress (Bruce 1977). Pierrehumbert and Beckman (1988) therefore suggest that the phrase accent is right-peripheral to the focused word rather than the phrase, and that it is attracted to the secondary stress from the word's periphery in compounds. Difficulties in the proposal aside, the idea of the phrase accent being the peak of a phrase is an important point in Pierrehumbert and Beckman's exposition of Swedish tone structure. In Bruce's work (1998), the same observation has been formulated as a strong expectation of a focal accent in each prosodic phrase. In Gussenhoven and Bruce (1999), the timing of the focal H (phrase accent) is described as related to the final stress of a word, i.e. the primary stress in non-compound words and the secondary stress in compound words. This solution is compatible with the characterization proposed by Pierrehumbert and Beckman (1988). In Riad (1998), a third solution is proposed. Using an optimality theoretic approach, Riad describes the association of the focal H in Stockholm Swedish (or

'prominence tone') by assuming a higher ranking of the constraint that associates the word accent (or 'lexical tone') than of the constraint that associates the focal H. The lexical tone has precedence over the prominence tone in associating to the syllable carrying primary stress, and the prominence tone is left to associate with the head of the following TBU (foot), namely the secondary stressed syllable.

Timing aside, the phrase accent's status as the peak of a phrase seems unproblematic, whereas the question of what type of phrase it signals is left unanswered. The fact that two phrase accents sometimes are found within one and the same prosodic phrase can be taken as evidence suggesting that the phrase accent is the head of a constituent smaller than the prosodic phrase, e.g. of an intermediate phrase. However, the lack of independent perceptual evidence supporting a division of these prosodic phrases into smaller phrases makes us reluctant to draw such a conclusion. Within ToBI (Beckman and Ayers 1993, Silverman *et al.* 1992), it is required that a break index of 3 (an intermediate phrase boundary) be used whenever a phrase accent label has been produced. However, as noted by Wightman (2002), this linkage between the break index tier and the tone tier de-emphasizes the perceptual experience by the listener. Here, we choose not to assume the existence of any constituents for which perceptual evidence cannot be found.

As regards a higher-level phonological constituent or category, Wightman *et al.* (1992) refer to a study by Liberman and Pierrehumbert (1984) where phonetic effects with a possibly larger domain than the intonational phrase were found although these phonetic effects have subsequently been argued to relate to discourse structure (Beckman and Pierrehumbert 1986). So far, we have taken the same approach when we described intonational coherence-signaling among prosodic phrases as a feature of discourse prosody in the previous chapter (chapter five). However, in the description of Swedish intonation, the 'prosodic utterance', a higher-level phonological constituent, is generally assumed. Thus, the prosodic hierarchy or tree assumed has a fixed depth with two levels of structure above the prosodic word. The elements of the prosodic utterance are exhaustively analyzed into a sequence of elements of the next-lower category, i.e. prosodic phrases and vice versa. All prosodic phrases leading up to the next strong boundary in the stream of speech are considered parts of one and the same prosodic utterance (as demanded by the Strict Layer Hypothesis).

Problems arise in transcription and description of spontaneous speech because of the expectation that a prosodic utterance is defined not only by the strong

boundaries delimiting it, but also by its internal prosodic structure, i.e. by some sort of coherence-signaling between the phrases in the utterance (such as phrasal downstep (van den Berg *et al.* 1992) or tonal coupling (Bruce 1982b)). Conversely, we also expect the prosodic phrases in sequences of downstepped/tonally coupled phrases – a prosodic utterance – to have prosodically weak endings utterance internally, which is not always the case.

In the latter case, when tonally coupled/downstepped phrases' endings are strongly marked, the definition of the prosodic utterance can be saved, by claiming that the strength of a boundary is not determined at the end of the unit it terminates but by the relation between the two units that it separates. The strongly marked ending is then presumably reevaluated by the listener and perceived as a weak boundary due to the following phrase's downstepped register. An example of such an utterance was given in the previous chapter (see chapter five, (5e) and Figures 5.10-5.12), and repeated here for the reader's convenience as (6a).

(6a)
$_1$där åt ju familjen inne i matsalen | $_2$och sen jungfrun hon fick nöja sig med att äta i köket | $_3$när dom hade ätit | $_4$så fick hon äta resterna | (*Bro_ow*)

$_1$the family ate in the dining-room | $_2$and then the maid had to be satisfied with eating in the kitchen| $_3$when they had eaten | $_4$then she got to eat the leftovers | (*Bro_ow*)

Respecting the demands of the Strict Layer Hypothesis and using the proposed distinction between two levels of structure above the prosodic word, the hierarchical structure of (6a) can be represented as in Figure 6.2.

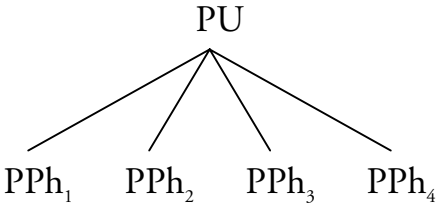


**Figure 6.2** *The hierarchical structure of (6a) respecting the demands of the Strict Layer Hypothesis. PPh is short for prosodic phrase, PU for prosodic utterance.*

This representation is nevertheless not entirely satisfactory since it resolves the conflict of criteria by simply ignoring certain cues. Ignoring cues weakens the independent phonetic evidence (such as relative pause length) upon which we would like the definitions of the two boundary types to rest. Furthermore, a lot of

information is lost if the prosodic structure of (6a) and the relations between the prosodic phrases contained therein is represented as in Figure 6.2. The difference in boundary strength between the first boundary and the following two is not acknowledged. The first boundary occurs between two prosodically coherent prosodic phrases, the second and third after prosodic phrases with strongly marked endings[6].

Two alternative hierarchies are given in Figure 6.3. Both structures violate the Strict Layer Hypothesis in that they contain an example of multiple domination ($PPh_3$); a phrase cannot have more than one mother node under the Strict Layer Hypothesis.

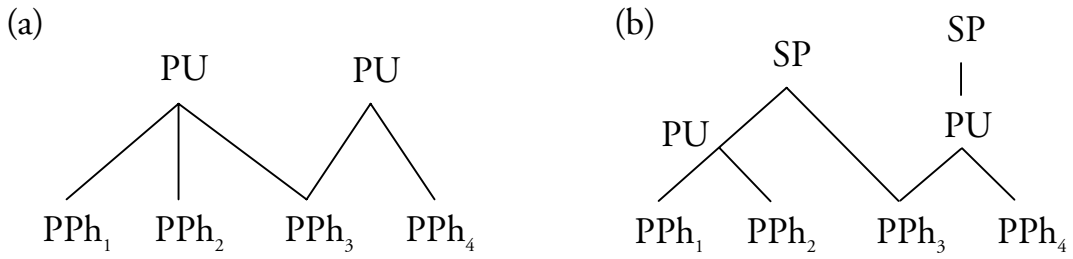


**Figure 6.3** *Alternative hierarchical structures of (6a) with two and three phrasal constituents respectively. SP is short for speech paragraph.*

Unlike in Figure 6.2, the close relation between the two first phrases is, however, recognized in Figure 6.3. In *b*, the difference in boundary strength between the first and second boundary – the boundaries after $PPh_2$ and $PPh_3$ – is also acknowledged. In order to make this distinction, a third constituent, the 'speech paragraph', has to be introduced into our hierarchy. Remember that differences in prosodic boundary

[6] For the reader's convenience, we repeat our prior analysis of (6a) in the following. The speaker talks about the situation of maids in Sweden in her youth: they had to eat in the kitchen while the family got to eat in the dining room. While the first prosodic phrase is separated from the following by a weak boundary, the end of the second phrase is more strongly marked prosodically. Nevertheless, the speaker chooses to make an addition (*när dom hade ätit* 'when they had eaten'). She makes clear that the maids had to wait to eat until after the family had eaten. The end of this prosodic phrase is also strongly marked prosodically (the new end of the sentence). The phrase is furthermore tonally coherent with the preceding. Thus, the addition is perceived as a continuation. The speaker then chooses to clarify the situation further, by adding the prosodic phrase *så fick hon äta resterna* 'then she got to eat the leftovers'. The temporal clause 'when they had eaten' can now be interpreted as either being a part of the preceding or the following sentence. Because the last clause cannot stand on its own syntactically, only one syntactic analysis is possible: the sentence boundary precedes the temporal clause. Prosodically, however, this analysis is not motivated: each prosodic phrase is a continuation of the preceding, and the temporal clause functions first to modify the first, and then the second sentence (or sentence-like unit).

strength are only possible under the Strict Layer Hypothesis if they reflect differences of boundary type.

Both structures in Figure 6.3 are unsatisfactory in that they weaken the independent phonetic motivation for assuming the different phrasal categories. Firstly, the multiple domination of $PPh_3$ makes it impossible to pinpoint the higher-level constituent boundary and consequently to use boundary strength in its definition. The structure in *b* furthermore makes use of a constituent – the 'speech paragraph' – that is not generally used in tonal transcription of Swedish. Strangert and Heldner (1995a and b) proposed that it be excluded also from the base prosody transcription system by collapsing what they term category 2 and 3 boundaries (i.e. utterance and paragraph boundaries) into one.

Possibly we are letting our knowledge of the syntactic structure of (6a) influence us in perceiving two higher-level constituents. Figure 6.4 illustrates another hierarchical representation where we return to the impression of (6a) as one sequence of prosodic phrases that are tonally linked.

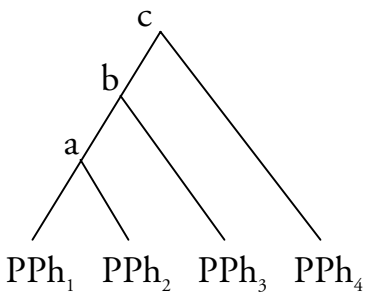


**Figure 6.4** *An alternative hierarchical structure of (6a).*

In the hierarchy in Figure 6.4 the close connection between the two first prosodic phrases is acknowledged, as is the fact that the second, third and fourth prosodic phrase are all continuations building on the preceding phrase and – at least for a while – signaled as new endings/as being terminal. However, this hierarchical representation is not possible under the Strict Layer Hypothesis either. We are assuming that prosodic phrases can be grouped into larger prosodic phrases (recursion), but that is not permitted. Alternatively, we may choose to assume three higher-level constituents (*a*, *b* and *c*) above the prosodic phrase and three boundary strengths in addition to the weak strength of the prosodic phrase boundary. Then we obey the demands of the Strict Layer Hypothesis, but intuitively we may feel that we are positing too many domain types. A consequence of representing the hierarchical structure of (6a) like this is also that moving from left to right in the

constituent *c*, the boundaries are expected to become increasingly stronger. They are not, and consequently this representation has no empirical support either.

In summary, we have intended to demonstrate that the assumption of a hierarchical structure of spontaneous speech above the prosodic phrase is problematic. There is no separate edge tone or tonal head to distinguish the prosodic utterance from the prosodic phrase, and the plan-as-you-go basis of spontaneous speech, makes the cues used to identify and distinguish between prosodic phrases and utterances in read speech ambiguous. Conflicts between boundary and coherence signaling cues have to be resolved by ignoring certain cues, e.g. differences in perceived boundary strength. Ignoring cues weakens the definability of the constituents assumed to the point where it seems justified to look for an alternative prosodic constituency. We would therefore like to propose that another way of accounting for differences in boundary strength than with differences of boundary types is sought for the description of spontaneous speech.

That another way of accounting for differences in boundary strength than with differences of boundary types is in fact needed, can be demonstrated if it can be shown that listeners exhibit good agreement in rating more than two boundary strengths above the prosodic word (e.g. continuous perception of boundary strength) (see Ladd 1996: section 6.3 for a discussion). Evidence to suggest that a large number of boundary strengths can be distinguished comes from a previous study on perceived boundary strength in Dutch (Sanderman 1996) where it was shown that listeners can rate boundary strength on a ten-point scale with good agreement. In the study described here, we did not ask listeners to judge boundary strength on a scale with a predetermined number of points, but instead used a continuous, so-called VAS scale. Such a scale does not force the listener to distinguish between more nor fewer degrees of boundary strength than (s)he finds adequate.

### 6.1.3 Research questions

In the study we performed, the goal was to relate perceived boundary strength to three known cues for prosodic phrasing in Swedish (pausing, F0 reset and final lengthening), and investigate whether the established division into two phrasal categories or constituents (the 'prosodic phrase' and the 'prosodic utterance') has empirical support in spontaneous Swedish. In order to examine these two issues, two perception experiments were carried out.

## 6.2 Experiment I: Method

### 6.2.1 Stimuli

In order to investigate the relationship between perceived boundary strength and the cues for prosodic phrasing, a total of 50 short speech fragments were chosen from the spontaneous part of the *SweDia 2000* database (Bruce *et al.* 1999). The speech database is described in more detail in section 3.2.1. Speech from five female and five male subjects, all from *Skåne*, was used (five speech sections from each speaker). Speakers from both generations were included.

All speech sections (typically one or two sentences long) contained at least one prosodic phrase boundary. In the perception experiment, the listeners were presented with an orthographic transcription of the spoken sentence(s) in which the boundary of interest was marked with a '/', as in (6b).

(6b) Sen tittar jag på lite såpor, / och så älskar jag musik.
'Then I watch some soaps, / and I love music.'

All boundaries to be judged by the listeners occurred either between complete sentences or between clauses, i.e. in syntactically motivated positions. In spontaneous speech, it cannot easily be determined whether two speech clauses are best described as two separate sentences introduced by discourse markers such as *and* or *but* or as two coordinated main clauses[7]. Therefore, no attempt was made to restrict the test to include only sentence or only clause boundaries.

We decided to give the listeners an orthographic transcription, in which the boundary of interest was punctuated either as a sentence boundary (i.e. with a full stop) or as a clause boundary (without a full stop and, in cases where appropriate, with a comma). We chose to do so expecting that the listeners, in particular those with a linguistic background, would make some sort of syntactic analysis and classification of the boundaries even if an orthographic transcription was not offered to them. By giving the listeners a transcription in which we guided them to listen to the boundary as either a clause or sentence boundary, we made an analysis of syntactic influence on listeners' perception of boundary strength possible. Due to the difficult classification of some boundaries in spontaneous speech discussed above, we would not have been able to make such an analysis otherwise. There

[7] Main clauses were only considered to be syntactically coordinated if they also demonstrated clear semantic coordination.

would have been no way of knowing whether the individual listener had perceived a given boundary as a sentence or clause boundary. That listeners indeed considered syntactic aspects of the stimuli, has been confirmed by listeners' comments after participating in the experiment. Some listeners also reported having reflected on the sentences' discourse relations (e.g. whether the two sentences on either side of the boundary of interest belonged to the same topic or not).

### 6.2.2 Listeners' task

The listeners were presented with a list of the 50 sentences/sentence pairs, and seated in front of a computer screen. Only the last word before the boundary of interest was displayed on the computer screen, see Figure 6.5 below. The subjects were instructed to read each sentence/sentence pair and listen to the recording of it. The recordings could be played and replayed by clicking on a button on the screen, and listeners could also go back to already given answers and change them. The subjects' task was to indicate how strongly marked they perceived the boundary in the recording. The whole experiment lasted roughly 20 minutes. The order of presentation of the stimuli was randomized and different for all listeners. A trial run containing three recordings preceded the actual test.

The instructions to the listeners were given on paper. After having read the instructions and made the trial run, subjects were given the possibility of asking questions to the experiment leader. No information concerning the task other than that occurring in the written instructions was given, however. The exact formulation of the task is given below in (6c).

(6c)
Din uppgift är att koncentrera dig på den markerade gränsen och avgöra hur starkt markerad den är i inspelningen. Upplever du en stark eller svag gräns, eller kanske inte någon gräns alls? Ditt svar anger du på datorskärmen genom att med musens hjälp placera markören någonstans mellan skalans vänstra (ingen gräns) och högra ände (mycket stark gräns).

'Your task is to concentrate on the marked boundary and determine how strongly marked it is in the recording. Do you perceive a strong or a weak boundary, or perhaps no boundary at all? You give your answer on the computer screen by using the mouse to place the marker somewhere between the scale's left (no boundary) and right end (very strong boundary).'

### 6.2.3 Visual analogue scale

The listeners indicated the boundaries' strength on a 'Visual Analogue Scale' (VAS). The VAS has been used in the measurement of clinical phenomena (such as pain) since the 1920s (Wewers and Lowe 1990), but it is suitable for measuring a variety of subjective non-clinical phenomena as well. Often the VAS is presented to subjects on paper. Its most common form is a 100 mm horizontal line. The subjects respond by placing a mark through the line at a position that represents their current perception of a given phenomenon (in this case a given boundary's strength) between the labeled extremes of the scale (in this case between *ingen gräns* 'no boundary' and *mycket stark gräns* 'very strong boundary', see Figure 6.5). The VAS is then scored by measuring the distance (in mm) from one end of the scale to the subject's mark on the line. The level of measurement represented by VAS data is usually assumed to be interval or ratio (Wewers and Lowe 1990).

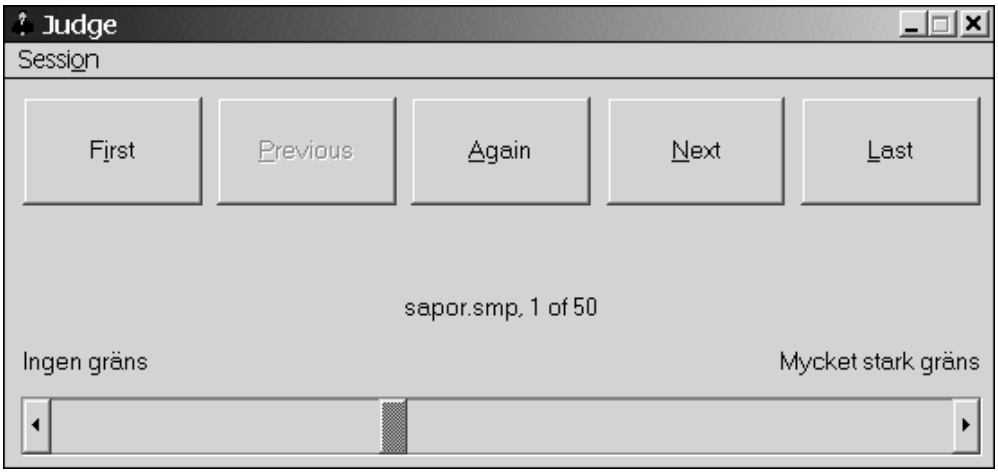


**Figure 6.5** *Screenshot of the computer program Judge (Granqvist 1996) where the listeners rated the strength of given boundaries by adjusting a scrollbar. The labeled extremes of the scale are:* ingen gräns *'no boundary' and* mycket stark gräns *'very strong boundary'.*

In our perception experiment, a computer-based VAS was used. The advantages and disadvantages of computer-based VAS scales are discussed in detail in Granqvist (1996). The main advantage and the reason for choosing a computer-based VAS in the present experiment, was the ease with which the scores are measured, and the possibility of letting the individual listener decide how many times (s)he wanted to listen to each stimulus. The main disadvantage of computer-VAS (Granqvist 1996) is the required familiarity with computers on part of the listener. In the trial run, we checked to make sure that the subjects understood how the computer program worked (that the stimuli could be replayed as many times as

the listener wished, and that already given answers could be changed), and that the sound volume was appropriate for the particular listener.

### 6.2.4 Listeners

A total of twenty subjects took part in the perception experiment. The subjects can be divided into four listener groups: experts, students with Swedish as native language, students with another native language than Swedish, and naïve listeners. The Swedish listeners represented speakers of different dialects of Swedish.

For the expert group, five phoneticians and other speech researchers (doctoral students and researchers) with experience from prosody research at a postgraduate level were engaged.

The student groups consisted of students in phonetics and general linguistics (no first-term students[8]) from the Department of Linguistics and Phonetics at Lund University, as well as doctoral students in general linguistics or a foreign language at Lund University. Five had Swedish as their native language, and five had a native language other than Swedish. All listeners in the student groups had some knowledge of prosody and prosodic transcription, but no experience of prosody research at a postgraduate level.

Finally, five naïve listeners without prior experience of prosodic transcription took part in the experiment.

The subjects were not paid for their participation, and the students received no course credits for participating.

### 6.2.5 Measurements and normalization

Three properties of the boundaries in the stimuli were investigated: the pause duration (in 33 of the 50 speech sections, the boundary was associated with a silent interval), the amount of final lengthening (or rather the difference in articulation rate between the penultimate and final word in the prosodic phrase preceding the boundary), and the reset of F0 across the boundary. Based on studies undertaken by House (1990) and House *et al.* (1998), we took the change in tonal levels at the boundary to be the most relevant F0 feature to measure. Although tonal movement configurations (perceived pitch fall on the final phrase element) were shown to have

[8] The course in prosody is given to the second-term students at the department.

impact on the percept of phrasing, it was concluded that the difference between tonal levels at the boundary provides the strongest boundary cue.

The duration of the pause may be perceived differently depending on whether the speaker talks fast or slowly. Therefore, in addition to measuring the actual duration of the pauses, the pause durations were also related to the speakers' speaking rates (the pause's length was divided by the average length of a syllable in the two prosodic words preceding the pause).

Since the inherent duration of a phone is known to be the largest source of variation in segmental duration, we would have liked to measure final lengthening as the difference between the duration of a given segment and the mean duration of that specific phone type. However, as discussed in section 3.1.4, in order to obtain the means of a speaker's phones, a fairly large amount of speech data needs to be segmented and labeled. As we had no possibility to do that, we chose to measure final lengthening in the same manner as in chapter three, i.e. we measured and compared the articulation rate in the two last words in the phrase. Since the measured articulation rate (number of syllables divided by the word's duration) is greatly affected by the size of the word (or rather the number of unstressed syllables it contains) all stimuli were chosen in such a way that the prosodic phrases before the boundaries of interest ended with two two-syllable prosodic words. The difference in articulation rate was normalized for differences in articulation rate between speakers and recordings by dividing it by the average articulation rate in the two words being compared.

The extent of F0 reset was measured in three different ways: 1) as the difference in F0 between the end point of the phrase before the phrase boundary of interest and starting point in the following phrase (measured in the last and first stable part of the phrases' F0 contours respectively), 2) as the difference in F0 between the end point of the phrase before the phrase boundary of interest and the first accent peak in the following phrase, and 3) as the difference in F0 between the last accent peak of the phrase before the phrase boundary of interest and the first accent peak in the following phrase (in Hz and semitones). We chose to use all three different ways of measuring F0 reset since each of the three ways has both advantages and disadvantages. The F0 end points are possibly more relevant to measure in order to get an insight into the perceived size of the F0 reset than the phrase-final accents. However, because of frequent phrase-final creak, they are not easily measured in a reliable way. The phrase-initial accent peaks are affected by the degree of

prominence assigned to the phrase-initial word, but the F0 starting point sometimes occurs in a stressed syllable and sometimes not.

## 6.3 Experiment I: Results and discussion

### 6.3.1 General characteristics of the boundaries in the stimuli

Two thirds of the examined boundaries (33 of 50) were associated with a pause (a silent interval). The pauses ranged in duration from 225 ms to 2847 ms. The F0 resets across the phrase boundaries[9] in the female subjects' speech ranged from a 185 Hz reset of F0 to a lowering of F0 by 4 Hz across the phrase boundary. The F0 resets across the phrase boundaries in the male subjects' speech ranged from a 51 Hz reset of F0 to a lowering of F0 by 15 Hz across the phrase boundary. The normalized difference in articulation rate between the phrase final and penultimate word ranged from 1.0 (corresponding to a change in articulation rate from 5.4 syllables per second to 1.8 syllables per second) to –0.3 (corresponding to a change from 4.1 syllables per second to 5.6 syllables per second).

None of the three variables investigated were strongly correlated with each other. There was no statistically significant correlation between the length of the pause and the change in articulation rate, nor between the extent of the F0 reset and the change in articulation rate. A weak positive correlation was, however, found between the extent of the F0 reset and pause length in the female subjects' speech ($r=.44$, $p<.05$, $n=25$) but only when F0 reset was measured as the difference in F0 (in semitones) between the start point and the end point (see section 6.2.5).

Apparently speakers do not, in general, maximize the use of several cues simultaneously to increase the strength of a boundary. This finding is partly consistent with results reported on in Horne *et al.* (1995) where only pause length was shown to conclusively be positively correlated with boundary strength. That study did not, however, investigate F0 reset as a cue to boundary strength.

### 6.3.2 The four listener groups

Statistical analyses revealed that the listeners agreed very well in their perceptual judgments. The Pearson correlation coefficient of each pairwise combination of

[9] The numbers reported here come from the measurements of F0 reset as the difference in F0 between the start and end point.

listeners in the four listener groups was significant at the .001 level. Following Sanderman (1996), we pooled the scores of the listeners for each boundary and calculated a mean to obtain an estimate of each boundary's perceived strength (in mm). First, we pooled the scores of the listeners in the different listener groups separately. A comparison of the scores of the different groups is presented below. It reveals very small differences between the groups, and consequently, in our analysis of the relationship between the cues to prosodic phrasing and perceived boundary strength in section 6.3.3, we calculated a PBS (Perceived Boundary Strength) score for each boundary using the scores of all listeners in all groups.

The PBS scores of the phoneticians and speech researchers (the expert listeners), ranged from 0 to 91 mm. The range of PBS scores assigned to the stimuli by the non-native students and naïve listeners was approximately the same (1-87 and 0-89 respectively). However, the students with Swedish as native language used a larger proportion of the scale. Their PBS scores ranged from 3 to 100. They also had a higher average PBS score than the other three listener groups, see Table 6.1.

Table 6.1 *Minimum, maximum, mean PBS scores and standard deviations for all four listener groups (in mm)*

| Listener group | Minimum PBS (mm) | Maximum PBS (mm) | Mean PBS (mm) | S.d. (mm) |
|---|---|---|---|---|
| Experts | 0 | 91 | 41 | 29 |
| Students (native) | 3 | 100 | 47 | 32 |
| Students (non-native) | 1 | 87 | 40 | 30 |
| Naïve listeners | 0 | 89 | 40 | 31 |

This variation between groups is, nevertheless, not large considering that each group was comprised of only five listeners. A single listener behaving differently from the others (e.g. with a strong bias towards assigning the boundaries high scores) can have a considerable influence on the group's PBS range and mean. The Pearson correlation coefficients of each pairwise combination of listener groups were also very high, as can be seen in Table 6.2.

Table 6.2 *R values of each pairwise combination of listener groups (p<.001)*

| | Experts | Students (native) | Students (nonnative) | Naïve listeners |
|---|---|---|---|---|
| Experts | 1.00 | .94 | .93 | .92 |
| Students (native) | | 1.00 | .96 | .95 |
| Students (nonnative) | | | 1.00 | .96 |
| Naïve listeners | | | | 1.00 |

The only obvious difference between the groups to which some importance needs to be attached, concerns the perceptual ratings of the boundaries not associated with pauses. If restricting our pairwise comparison of listeners' scores to only include the scores assigned to stimuli without pauses (n=17), then significance (at the .05 level) was reached primarily in the expert group. In the expert group, seven of the ten possible combinations are significant (although the *r* values are considerably smaller than in the experiment as a whole). In the student group (Swedish as native language), the Pearson correlation coefficient of one of the ten pairwise combinations reaches significance, and in the naïve and nonnative student group no combination reaches significance.

We interpret this finding as evidence preliminarily suggesting that pause length plays an important role in the percept of boundary strength, although possibly not quite to the same high degree for the expert listeners as for the other listeners participating in the experiment. In order to claim that boundary strength has communicative relevance, we should nevertheless perhaps not be primarily interested in trained listeners' (phoneticians') judgments.

### 6.3.3 Relationship between perceived boundary strength and pause length, F0 reset and final lengthening

In what follows, we will relate our acoustic measurements with the PBS scores calculated by pooling all listeners' scores (unless otherwise stated). Pooling all listeners' scores is motivated by the groups' similar percept of boundary strength. The perceived boundary strength (in mm) of the 50 stimuli ranges from 2 to 90, with a mean of 42 and a standard deviation of 30. In order to investigate a possible relationship between the three variables pause duration, extent of F0 reset and

change in articulation rate and perceived boundary strength, we will use multiple regression[10].

The correlation between pause length and perceived boundary strength proved to be very strong. Both the actual pause length (r=.92, $r^2$=.84, p<.001) and the normalized pause duration (r=.89, $r^2$=.79, p<.001) correlate strongly with perceived boundary strength. As shown in Figure 6.6, the longer the pause is, the stronger the listeners perceive the boundary. The only clear exceptions are pauses longer than 2 seconds, which are not perceived to be stronger than pauses of 1.5 to 2.0 seconds[11].

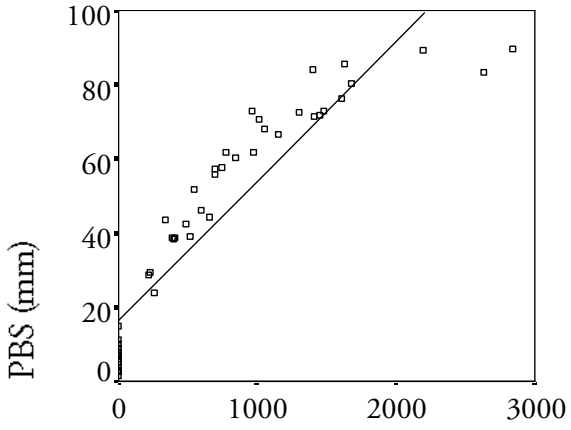


**Figure 6.6** *Scatterplot demonstrating the relationship between pause length and perceived boundary strength.*

However, 'change in articulation rate' proved not to be related to perceived boundary strength, and adding the variable 'extent of F0 reset' to the prediction of PBS scores made only a small unique contribution (r=.93, $r^2$=.87). Again, only when F0 reset was measured as the difference in F0 (in semitones) between the starting point and the end point (see section 6.2.5) did it make a contribution to the prediction of PBS scores.

We have so far not discussed the effect the number of cues used may have had upon listeners' perception of boundary strength in the stimuli. There are two reasons for this. Firstly, the experiment was not designed to test the effect of the number of cues upon perceived boundary strength. Consequently, each possible combination

[10] Multiple regression is a procedure that can determine which of the independent variables (cues to phrasing) best predicts or accounts for the performance on the dependent variable (PBS scores) or what combination of independent variables we need in order to account for the performance on the dependent variable.

[11] If these three outliers are excluded, the correlation coefficient *r* reaches a value of .96.

of cues (i.e. presence of all three cues, presence of phrase-final lengthening and pausing but not F0 reset, etc.) was not represented by the same number of stimuli in the test. Secondly, it is not a trivial matter to decide on a minimum amount of F0 reset and phrase-final lengthening needed in order to consider the cue in question present in a stimulus. Therefore, the only pattern that can be reported on here concerns the relationship between PBS and the presence vs absence of a pause. Boundaries associated with a pause are perceived as much stronger than those that do not involve any pausing. There are no clear indications of a relationship in the data between increasing PBS scores and an increased number of cues used by the speaker.

### 6.3.4 Relationship between perceived boundary strength and syntactic boundary type

There were two main types of boundaries that the listeners in the perception test were asked to judge as to boundary strength: boundaries punctuated as sentence boundaries and boundaries punctuated as clause boundaries. It is possible that the type of syntactic boundary affects the strength it is perceived to have. It is possible that boundaries transcribed as sentence boundaries (i.e. with a full stop in the orthographic transcription) were perceived as stronger than those transcribed as clause boundaries, due e.g. to listeners' expectations of stronger boundaries at sentence boundaries than clause boundaries.

Whether or not this is the case, cannot be tested by simply comparing the average perceived boundary strength of boundaries transcribed as sentence boundaries with the perceived boundary strength of boundaries transcribed as clause boundaries. A difference in perceived boundary strength resulting from such a comparison could easily reflect speakers' habits to mark sentence boundaries more strongly than clause boundaries, rather than an actual difference in how the same prosodic boundary is perceived depending on syntax. Indeed, the sentence boundaries in the stimuli are associated with both longer pauses and higher PBS scores than the clause boundaries, as shown in Table 6.3.

**Table 6.3** *Mean pause duration (in ms) and PBS score (in mm) for the different syntactic boundary types*

| | Syntactic boundary type | |
|---|---|---|
| | Sentence (n=26) | Clause (n=24) |
| Pause duration (ms) | 806 | 534 |
| PBS scores (mm) | 49 | 34 |

What we are interested in here is whether the relationship between pause duration and perceived boundary strength is the same regardless of syntactic boundary type, i.e. if a pause of a given duration is perceived as an equally strong boundary when it demarcates a sentence boundary and a clause boundary. In order to do so, we use linear regression[12]. The results are shown in Figure 6.7.

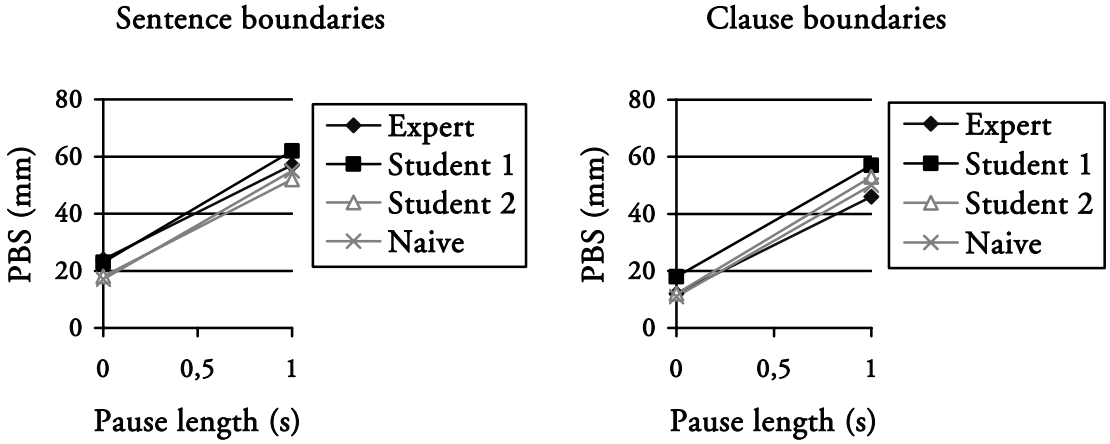


**Figure 6.7** *Graphs illustrating how PBS scores (in mm) increase with increasing pause length (in s). Student 1 is the listener group with Swedish students, student 2 is the listener group with non-native speakers of Swedish.*

As shown in Figure 6.7, the listeners assigned PBS scores of about 20 mm to sentences boundaries not associated with pauses. Clause boundaries not associated with pauses were assigned PBS scores of about 13 mm. Sentence boundaries were, in other words, perceived as more strongly marked than clause boundaries when they were not associated with pauses. This is true for all groups of listeners. The exact intercept values for each listener group is given in Table 6.4 (the constant *a*).

Regardless of syntactic boundary type, the PBS scores then increase with approximately 4 mm with each 100 ms prolongation of the pause (i.e. with 40 mm per second). This fact is reflected in the values of the *b* coefficient in Table 6.4. This means that a sentence boundary, regardless of its pause length, is assigned approximately 7 mm higher scores than a clause boundary with an associated pause of the same length.

The non-native students behaved somewhat differently than the other three listener groups in that the sentence boundaries' PBS scores increased more slowly (34 mm

[12] Linear regression brings out the relationship between two sets of values – which in this case is the relationship between the pause lengths and the PBS scores – and provides slope (the regression or *b* coefficient) and intercept values (the constant *a*).

per second or 3.4 mm per each 100 ms interval) with increasing pause length than the clause boundaries' PBS scores (4.1 mm per each 100 ms interval). We have no explanation for this fact, but due to the small number of subjects in each group, some random differences of this kind were expected (see section 6.3.2).

**Table 6.4** *The slope b (in mm per second) and intercept value a (in mm) of the relationships between pause length and PBS scores in the four listener groups*

| | Sentence boundaries | | Clause boundaries | |
|---|---|---|---|---|
| | a | b | a | b |
| Expert listeners | 24 | 33 | 12 | 34 |
| Student listeners, *native* | 23 | 39 | 18 | 39 |
| Student listeners, *non-native* | 18 | 34 | 12 | 41 |
| Naïve listeners | 17 | 38 | 11 | 39 |

It is clear that syntactic boundary type, in addition to the extent of the F0 reset, is a factor that affects perceived boundary strength (although either factor accounts for an equally large portion of the observed variation in PBS scores as pause length). A pause of a given length is not perceived as equally strong at a clause boundary and at a sentence boundary.

The fact that not only naïve listeners but also phoneticians' perception of boundary strength revealed a clear syntactic dependence, may indicate that the distinction made between prosodic phrases and prosodic utterances in studies of read speech is related to syntactic influence on perception.

### 6.3.5 Summary

The perception experiment reported on in the present chapter had two purposes. So far, we have only dealt with the first: the relationship between perceived boundary strength and three relevant cues to prosodic phrasing. The main finding is that pause length is the strongest cue for perceived boundary strength in the data studied. Boundaries not associated with pauses are perceived as weak or non-existent and listeners rarely agree in their perceptual judgments of them. The PBS scores of boundaries associated with pauses are, on the other hand, strongly correlated with the boundaries' pause length. The extent of F0 reset furthermore makes a minimal contribution to the prediction of PBS scores.

Another important finding is the influence made by syntax on the perception of boundary strength. A pause of a given length is perceived as a stronger boundary cue when it is found at a sentence boundary than at a clause boundary. Both phoneticians and listeners with less experience in prosodic transcription are influenced by syntax in their ratings.

The size of the change in articulation rate phrase-finally was not found to have any substantial effect on the perception of boundary strength, although the measurements of the change in articulation rate were admittedly crude.

A question that inevitably arises when analyzing the results of this perception experiment, is if the different measurements made of F0 reset and final lengthening somehow have failed to capture these variables' importance for perceived boundary strength. Other ways of measuring and normalizing the three examined variables may need to be sought. The measurements of F0 reset and final lengthening used are admittedly not comparable with those made under ideal conditions in studies of read speech. We therefore hesitate to claim that these variables are unimportant for the perception of boundary strength in spontaneous speech based only on the results of this perception experiment. Their weak effect on perceived boundary strength in spontaneous speech may be a consequence of an unsystematic relation in spontaneous speech between the size of the F0 reset and the change in articulation rate on the one hand and the pause length on the other. More controlled studies of e.g. elicited speech are warranted to examine this issue further.

The unexpectedly weak influence of variables other than pause length on perceived boundary strength and the necessary crudeness of our acoustic measurements, also motivate further studies into the role of pausing for the percept of boundary strength. The fact that pauses of varying length also occur frequently elsewhere in spontaneous speech, not as signals of prosodic phrase boundaries, may also make us reluctant to place too much importance on their presence and length.

Based on a suggestion made by Jan-Olof Svantesson at one of the department's Friday seminars, I therefore decided to run the test a second time, this time with a fixed pause length in the stimuli. The purpose was still to investigate the relationship between extent of F0 reset and change in articulation rate on the one hand and perceived boundary strength on the other, but in this case by controlling for pause length.

## 6.4 Experiment II: Method

In the second perception experiment, all pauses in the stimuli were manipulated so as to have the same length. The idea was to test if pause length co-varies with other perceptually relevant correlates to boundary strength. If the results of the second perception experiment were to reveal that listeners are able to agree in their perceptual judgments of the stimuli with manipulated pause lengths, then we would interpret this as evidence supporting the idea that pause length is correlated with other perceptually relevant correlates to boundary strength, cues that we nevertheless have failed to capture in our acoustic measurements. However, if the listeners were not able to agree in their perceptual judgments, then we would feel confident in concluding that pause length is the main correlate of boundary strength in the spontaneous speech data examined.

Due to the manipulation of the pauses' lengths, we did not expect the listeners to necessarily assign the boundaries in the second experiment the same scores as in the first experiment. If pause length were correlated with one or several other cues to boundary strength, we would nevertheless expect the boundaries to be rank-ordered in the same fashion in both experiments (boundaries associated with longer pauses than in the first run could perhaps obtain higher scores, and boundaries associated with shorter pauses than in the first run could perhaps be assigned lower scores). If the listeners in Experiment II were to rank-order the stimuli in the same manner as in experiment I, then we would return to the acoustic registrations of the production data and make further measurements of possible correlates to boundary strength. However, if the listeners were to perceive all boundaries as equally strong or only to demonstrate a random pattern in their perceptual judgments in Experiment II, then we would conclude that pause length is the main correlate to perceived boundary strength in the spontaneous speech data examined.

Note that such a result would not mean that the other correlates under investigation are not relevant for the perception of a prosodic phrase boundary in spontaneous speech. It would merely suggest that perceived boundary *strength* is determined by pause length.

### 6.4.1 Stimuli

Out of the total of 50 short speech sections that were used in the first perception experiment (see section 6.2.1), 33 were selected for inclusion in the second experiment; all stimuli where the syntactic boundary of interest was associated with

a pause. Inserting pauses in boundaries where no pause was made in the original recording proved unsuccessful. The resulting manipulated speech sounded too unnatural for the stimuli to be used in the perception experiment. Shortening and prolongation of existing pauses could, thanks to the high quality of the recordings (in particular the minimal background noise), be made successfully.

The pauses associated with the boundaries in the subset of stimuli that were chosen for inclusion in the second experiment were set to 800 ms. A 800 ms long pause represents a pause that sounds perfectly natural as a boundary cue both at sentence and clause boundaries (see the means in Table 6.3).

### 6.4.2 Listeners and task

The listeners' task was the same as in Experiment I (see section 6.2.2). They were presented with a list of the 33 sentences/sentence pairs, and seated in front of a computer screen. The last word before the boundary of interest was displayed on the computer screen, and a complete orthographic transcription of each stimulus was given on paper. The subjects were instructed to read each sentence/sentence pair and listen to the recording of it. The recordings could be played and replayed by clicking on a button on the screen. The subjects' task was to indicate how strongly marked they perceived the boundary in the recording on a so-called Visual Analogue Scale (VAS) (see section 6.2.3). The whole experiment lasted roughly 15 minutes. The order of presentation of the utterances was randomized and different for all listeners. A trial run containing three recordings preceded the actual test.

Fifteen subjects participated in the second perception experiment. The same kinds of listener groups that participated in Experiment I (see section 6.2.4) were asked to participate in Experiment II: five phoneticians and speech researchers with experience of prosody research, five students (all with some knowledge of prosodic labeling) and five naïve listeners. All students participating in the second experiment had Swedish as their native language.

The individual listeners were not the same as those participating in the first experiment. The second experiment was designed after the first experiment had been run, and therefore we did not have the choice to e.g. divide our listeners into two groups with one taking Experiment I first and the other Experiment II.

## 6.5 Experiment II: Results and discussion

None of the Pearson correlation coefficients obtained from comparisons of each pairwise combination of listeners in the three listener groups was significant ($p>.05$). The listeners did not agree in their perceptual judgments. Furthermore, the listeners used a considerably smaller part of the scale in Experiment II than in Experiment I. The expert listeners' PBS scores (in mm) range from 21 to 66, the student listeners' from 20 to 58 and the naïve listeners' from 26 to 76, see Table 6.5. The PBS scores of all listeners pooled, range from 24 to 58 (mean=41, s.d.=8). This is to be compared with the PBS scores in Experiment I, which for the three listener groups in question (expert, students with Swedish as their native language and naïve listeners) and the 33 stimuli used in both experiments, ranged from 19 to 93 (mean=61, s.d.=20).

**Table 6.5** *Minimum, maximum, mean PBS scores and standard deviations for all three listeners groups (in mm)*

| Listener group | Minimum PBS (mm) | Maximum PBS (mm) | Mean PBS (mm) | S.d. (mm) |
|---|---|---|---|---|
| Experts | 21 | 66 | 46 | 12 |
| Students (native) | 20 | 58 | 35 | 8 |
| Naïve listeners | 26 | 76 | 45 | 14 |

Poor agreement in perceptual judgments and/or little variation in perceived boundary strength between the stimuli were both results we expected if pause length indeed is the main correlate to boundary strength in spontaneous speech (see section 6.4). Since the pause lengths in the stimuli had been manipulated, we did not expect the stimuli's PBS scores to be identical in the two experiments even if there were other cues in the speech signals besides pause length that signal boundary strength. It was to be expected that boundaries with pauses that had been reduced in duration would be perceived as weaker, and boundaries with pauses that had been increased in duration would be perceived as stronger. However, the rank-ordering of the stimuli was still expected to be the same. In Tables 6.6a and b, we have rank-ordered the stimuli according to their PBS scores (from the stimulus with the lowest PBS score to the stimulus with the highest PBS score). To facilitate the comparison, only those 33 stimuli that were used in both experiments have been rank-ordered.

**Table 6.6a** *The relative rankings of the stimuli that were used in both experiments (I and II): ranks 1 through 17. The stimuli are ranked from the weakest to the most strongly perceived boundary. Numbers in parentheses are PBS scores (in mm).*

| Rank | All listeners[13] | | Experts | | Swedish students | | Naïve | |
|---|---|---|---|---|---|---|---|---|
| | I | II | I | II | I | II | I | II |
| 1 | Ämnen (19) | Ämnen (24) | Ämnen (12) | Skolan (21) | Ämnen (23) | Frukost (20) | Saker (19) | Ämnen (26) |
| 2 | Kurser (29) | Året (26) | Ensam (31) | Åtta (22) | Kurser (31) | Året (22) | Ämnen (23) | Innan (27) |
| 3 | Ensam (31) | Skolan (26) | Filtar (31) | Filtar (25) | Ensam (34) | Vecka (25) | Kurser (26) | Vecka (29) |
| 4 | Filtar (37) | Åtta (34) | Skolan (31) | Ämnen (28) | Saker (39) | Gånger (26) | Ensam (28) | Året (30) |
| 5 | Linje (37) | Honom (34) | Kurser (32) | Året (30) | Mycket2 (41) | Länder (26) | Mycket2 (28) | Fria (32) |
| 6 | Saker (37) | Vecka (34) | Linje (36) | Kor (35) | Drömmar (43) | Fria (27) | Göra (31) | Honom (33) |
| 7 | Göra (41) | Kor (35) | Honom (39) | Kronor (35) | Linje (43) | Villa (27) | Linje (33) | Skolan (33) |
| 8 | Drömmar (43) | Ensam (37) | Göra (43) | Saker (38) | Nånting (44) | Ämnen (28) | Filtar (36) | Dessa (34) |
| 9 | Mycket2 (44) | Fria (37) | Drömmar (47) | Ljuset (40) | Filtar (45) | Dessa (29) | Drömmar (38) | Ensam (35) |
| 10 | Skolan (44) | Kurser (37) | Mycket1 (50) | Såpor (40) | Göra (48) | Drömmar (29) | Också (41) | Drömmar (37) |
| 11 | Honom (48) | Länder (37) | Saker (52) | Vecka (40) | Honom (51) | Kurser (30) | Skolan (44) | Hemma (37) |
| 12 | Nånting (52) | Saker (37) | Gånger (53) | Honom (43) | Skolan (56) | Skolan (30) | Såpor (46) | Mycket1 (37) |
| 13 | Också (58) | Hemma (38) | Drängar (54) | Ensam (44) | Vecka (64) | Ljuset (31) | Åtta (50) | Göra (38) |
| 14 | Åtta (59) | Mycket1 (38) | Åtta (55) | Fria (44) | Dessa (67) | Åtta (33) | Honom (54) | Kor (39) |
| 15 | Dessa (61) | Filtar (40) | Frukost (56) | Gånger (44) | Frukost (69) | Hemma (33) | Dessa (55) | Nånting (39) |
| 16 | Mycket1 (63) | Linje (40) | Nånting (56) | Länder (45) | Gånger (69) | Mycket2 (33) | Nånting (57) | Villa (39) |
| 17 | Gånger (63) | Ljuset (40) | Året (57) | Frukost (46) | Också (69) | Ridning (33) | Jobbet (64) | Kurser (42) |

[13] Non-native students are not included since they only participated in the first experiment.

**Table 6.6b** *The relative rankings of the stimuli that were used in both experiments (I and II): ranks 17 through 33. The stimuli are ranked from the weakest to the most strongly perceived boundary. Numbers in parentheses are PBS scores (in mm).*

| Rank | All listeners | | Experts | | Swedish students | | Naïve | |
|---|---|---|---|---|---|---|---|---|
| | I | II | I | II | I | II | I | II |
| 17 | Gånger (63) | Ljuset (40) | Året (57) | Frukost (46) | Också (69) | Ridning (33) | Jobbet (64) | Kurser (42) |
| 18 | Såpor (64) | Drömmar (41) | Dessa (60) | Linje (47) | Åtta (73) | Jobbet (34) | Mycket1 (64) | Saker (43) |
| 19 | Året (67) | Villa (41) | Hemma (60) | Kurser (48) | Mycket1 (74) | Mycket1 (35) | Gånger (68) | Åtta (45) |
| 20 | Jobbet (69) | Innan (42) | Mycket2 (63) | Mycket (49) | Jobbet (76) | Drängar (37) | Året (69) | Ridning (45) |
| 21 | Drängar (71) | Dessa (44) | Kor (64) | Nånting (51) | Året (77) | Honom (37) | Kor (70) | Länder (46) |
| 22 | Frukost (71) | Frukost (44) | Också (64) | Hemma (54) | Villa (77) | Kor (37) | Vecka (71) | Ljuset (47) |
| 23 | Vecka (71) | Göra (44) | Såpor (64) | Innan (54) | Ljuset (80) | Linje (39) | Drängar (73) | Linje (48) |
| 24 | Kor (72) | Gånger (45) | Jobbet (67) | Villa (54) | Kor (81) | Saker (39) | Kronor (73) | Såpor (49) |
| 25 | Hemma (74) | Ridning (45) | Kronor (70) | Göra (55) | Såpor (82) | Också (40) | Hemma (74) | Drängar (54) |
| 26 | Kronor (76) | Såpor (46) | Länder (73) | Jobbet (58) | Drängar (85) | Filtar (41) | Länder (79) | Gånger (55) |
| 27 | Länder (79) | Nånting (48) | Vecka (77) | Mycket2 (59) | Länder (85) | Ensam (42) | Kriget (80) | Jobbet (61) |
| 28 | Villa (81) | Jobbet (49) | Kriget (80) | Drömmar (60) | Kronor (86) | Göra (42) | Ljuset (84) | Kriget (62) |
| 29 | Ljuset (83) | Drängar (51) | Villa (82) | Också (60) | Hemma (89) | Såpor (45) | Villa (85) | Filtar (63) |
| 30 | Kriget (85) | Kronor (55) | Innan (83) | Dessa (62) | Kriget (93) | Innan (46) | Fria (87) | Mycket2 (66) |
| 31 | Innan (88) | Mycket2 (55) | Ljuset (83) | Drängar (62) | Innan (94) | Nånting (46) | Innan (87) | Också (69) |
| 32 | Ridning (90) | Också (57) | Ridning (85) | Kriget (62) | Ridning (97) | Kronor (47) | Frukost (89) | Frukost (73) |
| 33 | Fria (93) | Kriget (58) | Fria (91) | Ridning (66) | Fria (100) | Kriget (58) | Ridning (89) | Kronor (76) |

As can be seen in Tables 6.6a and b, the rank-order of the stimuli was different in the two experiments. All listener groups rank-ordered the stimuli in a different way when the pause length was held constant. Pause length clearly plays an important role in the percept of boundary strength, although – once again – not to the same high degree for the expert listeners as the other listeners. The rank-ordering of the stimuli in Experiment I on the other hand, differs little between the groups. The stimulus *Ämnen*, e.g., was judged to have the weakest boundary by the experts and Swedish students, while the naïve listeners ranked it as having the second weakest boundary, only 4 mm stronger than the stimulus they ranked as having the weakest boundary. Similarily, the stimulus *Fria* was ranked as being associated with the strongest boundary by the experts and the Swedish students, and as being associated with the second strongest boundary by the naïve listeners (only 2 mm weaker than the strongest stimulus).

A few stimuli were assigned similar ranks in the two experiments. The stimulus *Ämnen*, e.g., was judged to have the weakest boundary in both Experiment I and II (all listeners). Similarily, the stimulus *Kriget* was ranked as being associated with the fourth strongest boundary in Experiment I, and as being associated with the strongest boundary in Experiment II (all listeners). The boundary in *Ämnen* (see (6d)) is special in that it occurs in a disfluent utterance, as is the boundary in *Kriget* (see (6e)) since it is preceded by a considerable reduction in articulation rate (from 5.4 in the penultimate word to 1.8 syllables per second in the final word).

(6d)
Just nu studerar jag på Komvux (p) för att läsa in (p) några ämnen / som jag inte (p) läste på gymnasiet innan.

'Right now I'm studying at Komvux (p) taking some (p) courses / that I didn't (p) take in high-school'

(6e)
Och det var under kriget. / Jag gick i skolan ju.

'And that was during the War. / I went to school then.'

Since many stimuli obtained the same rank (i.e. were assigned the same PBS score), but not equally many in all listener groups or both experiments, a visual comparison between the rank-ordering of the stimuli in the different columns of Tables 6.6a and b is not straightforward. However, the two rank-orders (the rank-order of stimuli in Experiment I and the rank-order of stimuli in Experiment II)

can be related using a Spearman ρ. The Spearman Rho correlation coefficients (ρ) of the correlation between Experiment I and II are non-significant in all groups ($p>.05$). No group of listeners rank-ordered the stimuli in the same manner in both experiments. There are no two groups that rank-ordered the stimuli in a simular manner in Experiment II ($p>.05$) either. No two listener groups were able to agree on another correlate to boundary strength and make consistent use of it in their perceptual judgments of boundary strength in Experiment II.

In summary, if there are cues that we have failed to measure but that correlate with pause length and therefore also with the perceived strength of boundaries, then they are generally not strong enough to make a real difference when they no longer co-vary with pause length (see, however, (6d) and (6e)).

### 6.5.1 Relationship between perceived boundary strength and syntactic boundary type

The only factor that had a systematic effect on the perceived boundary strength in the second perception experiment was syntactic boundary type.

There were two main types of syntactic boundaries that the listeners in the perception test were asked to judge as to boundary strength: sentence boundaries and clause boundaries. The results of experiment I indicated that sentence boundaries, regardless of pause length, are assigned some 7 mm higher scores than clause boundaries. The effect of syntactic boundary type on perceived boundary strength could also be observed in the responses given by the listeners in Experiment II.

In contrast to the first perception experiment, we could test syntactic boundary type's effect on perceived boundary strength by simply comparing the average perceived boundary strength of boundaries transcribed as sentence boundaries ($n=14$) with the perceived boundary strength of boundaries transcribed as clause boundaries ($n=19$). A difference in perceived boundary strength resulting from such a comparison would in this case, where pauses were equal in length in both groups, reflect a difference in how one and the same boundary (or rather pause length) is perceived depending on syntax. There was a difference in PBS scores between sentence and clause boundaries even in Experiment II. Sentence boundaries (45 mm) were given some 8 mm higher PBS scores than clause boundaries (37 mm), which is comparable to the difference in observed PBS scores in Experiment I.

### 6.5.2 Summary

The purpose of the second perception experiment was to test whether pause length co-varies with other perceptually relevant correlates to boundary strength. The finding that listeners could no longer agree in their perceptual judgments of boundary strength when pause length was held constant, and that all listener groups rank-ordered the boundaries' strength in a different fashion than when the pauses in the stimuli had not been manipulated, is interpreted as supporting pause length as the main correlate to boundary strength in our spontaneous speech data, as is the fact that the variation in PBS scores was smaller in Experiment II than I.

If there are cues that correlate with pause length and therefore also with the perceived strength of boundaries in spontaneous speech (but which we have not been able to measure in our production data), then they are at least not strong enough to make a real difference when they no longer co-vary with pause length.

## 6.6 Discussion

### 6.6.1 Implications for the number of phrasal categories in spontaneous Swedish

The perception experiments reported on in the present chapter had two purposes. Firstly, to determine what the relationship between perceived boundary strength and three known cues to prosodic phrasing in Swedish is. Although our spontaneous speech material has allowed us to make only rather crude acoustic measurements, a vast effect of pause length on perceived boundary strength has been shown to exist. Furthermore, extent of F0 reset and syntax also appear to have some effect upon perceived boundary strength whereas amount of final lengthening (change in articulation rate phrase-finally) would appear to matter little for the percept of boundary strength. Unlike the influence of syntax, which is small in comparison to pause length, the effect of pause length on perceived boundary strength is of a gradient kind. As pause length increases gradiently, the perceived boundary strength also increases gradiently or continuously. The second purpose of the perception experiments in this chapter, was to investigate whether the established division into two phrasal categories or constituents (the 'prosodic phrase' and the 'prosodic utterance') has empirical support in spontaneous Swedish.

The results of the perception experiments demonstrate that listeners' perception of boundary strength in spontaneous speech is continuous rather than categorical.

Listeners have no difficulties in judging the strength of boundaries with high agreement on a continuous scale. This finding does not support a categorical distinction between weak/phrase boundaries and strong/utterance boundaries, nor does the apparent lack of existing evidence to support a communicatively meaningful distinction between strong and weak boundaries. Unlike the distinction between three degrees of prominence in the modeling of Swedish, which rests on convincing empirical evidence, it is difficult to think of a minimal pair where the meaning would change as a weak, prosodic phrase boundary is changed into a strong, prosodic utterance boundary. On the other hand, examples that show that the presence or absence of a prosodic phrase boundary cause a change of meaning are not difficult to find, see (6f). Removing the prosodic phrase boundary in 1 leads to a change of meaning (see 2), whereas changing the strength of the boundary (compare 2 and 3) does not.

(6f) $_1$Jag vill inte. | Kasta den.
$_2$Jag vill inte kasta den.
$_3$Jag vill inte. || Kasta den.

$_1$'I don't want to. | Throw it away.'
$_2$'I don't want to throw it away.'
$_3$'I don't want to. || Throw it away.'

(Examples from Bruce 1998: 17)

A minimal pair with meanings that possibly change as the weak, prosodic phrase boundary is changed into a strong, prosodic utterance boundary has been suggested to me by Merle Horne (personal communication), see (6g).

(6g) $_1$Jag var med om en intressant händelse: | män berätta(de) om kärlek.
$_2$Jag var med om en intressant händelse. || Men berätta om kärlek.

$_1$'Something interesting happened to me: | men talked about love.'
$_2$'Something interesting happened to me. || But tell me about love.'

A similar minimal pair (between *långa män* 'tall men' and *långa, men* 'tall but') has been tested perceptually by Gårding and Eriksson (1989). In that study, it was nevertheless concluded that the perceptually relevant difference was one of accentuation. Accentuation of *män/men* [mɛnː] led to the association of the utterance with the meaning 'tall men', whereas deaccentuation of [mɛnː] led to the association of the utterance with the meaning 'tall but'.

In descriptions of spontaneous speech, the usefulness of the 'prosodic utterance' as a phonological category in the prosodic constituency can also be questioned based on the observation that sequences of prosodic phrases separated by weak boundaries, i.e. per definition prosodic utterances, do far from always demonstrate an internal prosodic structure like that resulting from phrasal downstep/tonal coupling. This finding is similar to the findings that led Beckman and Pierrehumbert (1986) to analyze declination and final lowering as controlled by discourse structure rather than as phonetic effects indicating the presence of a higher-level constituent such as the prosodic utterance. In the evaluation of the base prosody system (Strangert and Heldner 1995a and b), it was proposed that the definitions of the different prosodic categories be refined as one felt that the existing system to some extent is unclear. We agree in that the definitions of the prosodic groups above the prosodic word are unclear, but believe that the difficulties associated with defining the difference between the prosodic phrase and the prosodic utterance in spontaneous speech is related to the continuous (as opposed to categorical) nature of the perception of boundary strength. As there is no separate edge tone or tonal head to distinguish the prosodic utterance from the prosodic phrase, the definitions of the two constituents rest on listeners' ability to distinguish between weak and strong boundaries. The continuously varying boundary strength in (both the production and perception of) spontaneous speech makes this distinction difficult to make and the definition unsatisfactory.

### 6.6.2 Phrasal structure in spontaneous Swedish

We would like to propose that degree of boundary strength at this level of structure be described as degree of emphasis is described in Swedish. Although perceptible differences can be produced and perceived, the differences of degree can be described as variation within the same phonological category[14] (in the case of emphasis, within the third degree of prominence 'focus', see section 1.4). Hirschberg and Pierrehumbert (1986) have suggested the same for English. They propose that pause duration and the extent of F0 reset at phrase boundaries are manipulated gradiently in speech production as a reflection of discourse structure. The tonal coherence that sometimes can be observed among prosodic phrases in Swedish can also be analyzed as related to discourse structure rather than indicating the existence of a higher-level phonological unit. The units within which phrasal downstep/tonal coupling operate (see chapter five), i.e. the units within which we

[14] This view is similar to that of Ladd and Morton (1997) who show that the distinction between 'normal' and 'emphatic' accents in English is not categorically perceived.

have found tonal coherence to be produced and perceived, are above the clause level, and often also above the sentence level, i.e. per definition discourse units (Stubbs 1983).

What we believe to be important to pay attention to in future analyses of discourse prosody in spontaneous speech is the global signaling of continuity across phrase boundaries, i.e. whether or not a given chunk of speech, 'prosodic phrase', is prosodically linked to the preceding. The prosodic backward linking appears to be more communicatively important than the local signaling of terminality at phrase endings in spontaneous speech. As discussed in the previous chapter (section 5.3.2), the possibility of overriding prosodic signals of terminality and boundary strength (as mainly signaled by pause length) with coherence signaling cues is an important asset to the speaker in spontaneous speech. The disambiguation of a structure such as that in (6h) (i.e. understanding that the preposition phrase *på datan* 'on the computer' belongs to the preceding rather than the upcoming speech), is not to be thought of as depending on boundary strength (i.e. pause length in spontaneous speech), but on the presence of tonal coherence over the boundary, i.e. a more global phenomenon than boundary strength (example from the spontaneous recording of *Bro_ym*).

(6h)
det första jag går upp och tittar det är mina mail (1.54 s pause) på datan (*Bro_ym*)

'the first thing I go up to check is my [e-]mail (1.54 s pause) on the computer' (*Bro_ym*)

Whereas degree of boundary strength would appear to co-vary with the degree of perceived coherence across the boundary in read speech (and motivate a categorical distinction between two phrasal constituents, definable in terms of boundary strength), the situation in spontaneous speech is clearly more complex. Tonal coherence in spontaneous speech can be perceived across boundaries that are strongly marked locally. Nevertheless, we choose not to redefine boundary strength as being determined by the global signaling of tonal coherence, and therefore not to distinguish between prosodic phrase boundaries and prosodic utterance boundaries in our description of spontaneous speech by referring to the degree of tonal coherence perceived across the boundary. Such a redefinition of boundary strength has no support in our perceptual data. Instead, we analyze tonal coherence as controlled by discourse structure in spontaneous speech, and note that local boundary signals vary in such a way that they cannot be analyzed as indicating the presence of a higher-level constituent in spontaneous speech.

CHAPTER 7

# Summary

The purpose of this study was to investigate how prosody is used to divide the flow of speech into chunks appropriate in size for production and perception of spontaneous speech, i.e. appropriate for the speaker to plan and produce in the spontaneous speech situation, and appropriate for a successful conveyance of the message to the listener. Thereby, the study represents a move away from the laboratory speech examined in many previous related studies. Although a Southern Swedish speech material has been examined, the study is not intended as a study of the Southern Swedish dialect (*skånska* 'Scanian'); rather we have used Southern Swedish as a convenient object on which to test various hypotheses about the phrasing function of prosody in spontaneous speech. The hypotheses were formulated to test the relevance of findings made in previous production and perception experiments with read Swedish speech materials, predominantly with recordings of the so-called standard variety of Swedish. The issues dealt with in the study concern both the phonetics and the phonology of prosodic phrasing in spontaneous Swedish.

After a general introduction to prosodic phrasing in chapter one, an attempt was made to determine whether empirical evidence can be found for theoretical claims made about a number of universal optimality theoretic constraints on prosodic

phrasing in chapter two. The distribution of prosodic phrase boundaries in spontaneous speech was investigated by regarding it as the reflection of constraints that restrain the production and perception of speech (e.g. constraints on the amount of speech that can be produced without making a planning stop and constraints on how the speech can be chunked up in relation to syntactic and information structure without rendering the speech incomprehensible to the listener).

Evidence of a cohesional strategy, a Wrap-XP constraint, was found whereby a maximal projection is wrapped into a single prosodic phrase, in addition to traces of constraints on the prosodic phrases' maximal and minimal size. Furthermore, evidence against a high ranking of the constraint Align-Focus,R, a constraint that right-aligns a focus constituent in information structure with a phrase in prosodic structure, was found which was related to the non-peripheral nature of phrase accents in Swedish. However, a strategy by which the prosodic marking of the focal information is reinforced by left-aligning it with a prosodic phrase boundary was observed, i.e. evidence of an Align-Focus,L constraint. Examination of so-called disfluencies or speech repairs, a common feature of spontaneous speech, revealed two different strategies: a cohesional strategy across 'abridged repairs' and 'modification repairs', and a demarcative strategy helping listeners to identify 'fresh starts'. Differences in the realization between phrase boundaries at fresh starts and other positions, suggested a division of the Align-XP constraint into two: Align-XP,L and Align-XP,R.

Chapter three presented results from an examination of the phonetic signaling of boundaries in spontaneous speech.

In an analysis of changes in articulation rate within the prosodic phrase (measured in syllables per second in each prosodic word of the phrase), a significant effect of position was found. Phrase-initial words were more quickly articulated than phrase-internal words, and phrase-final words were found to be more slowly articulated than both phrase-initial and phrase-internal words. In 3-word phrases, a progressive articulation rate reduction could consequently be observed, but the analysis of 4- and 5-word phrases revealed that significant differences in articulation rate between successive words are generally confined to the final part of the prosodic phrase (between the penultimate and final word). The perceptual reality of the measured fast articulation rate in phrase-initial words was discussed. In order to answer the question of whether or not such a phenomenon as phrase-initial shortening exists in Swedish, it was concluded that further investigations are needed. It was, however,

concluded that phrase-final lengthening, as observed by a difference in articulation rate between phrase-final and penultimate words, exists in Southern Swedish despite the dialect's prosodic similarities with Danish, e.g. the lack of a (high) phrase accent. Previous studies have related phrase-final lengthening, which exists in Swedish but not in Danish, to the Stockholm Swedish phrase accent and its rise-fall F0 gesture. Further support for the hypothesis that final lengthening is a learned and language-specific phenomenon has thereby been found. A rank ordering of the data furthermore revealed that final lengthening had been used in about 4 of 5 phrases, suggesting that duration features are just as important cues to prosodic phrase structure in spontaneous speech as they have been reported to be in read speech.

In chapter four, the tonal means used in speech to signal coherence within the prosodic phrase were investigated.

By examining the relationships between F0 slope, phrase length and F0 starting point (the F0 value at the beginning of the prosodic phrase's F0 contour and/or in the first accent peak), we made an attempt to test the two Lund intonation models' capacities for describing spontaneous speech. The success with which spontaneous speech can be described within the models is interesting as the two approaches have different implications for the amount of preplanning needed. The original Lund model is characterized by assuming length-dependency, and thereby a requirement of advance planning of the whole phrase, whereas the revised Lund model assumes no such hard preplanning. Indications of lookahead were not found in the data. Neither was there evidence to suggest that speakers vary F0 slope to accommodate for differences in phrase length in the data, nor signs to suggest that speakers begin long phrases with higher F0 than short phrases. As predicted by the revised Lund model, a relationship between F0 starting point and F0 slope was nevertheless found. Consequently, it was concluded that the internal structure of prosodic phrases in spontaneous speech is most successfully described within the revised Lund model.

In chapter five, we followed up on the finding that F0 starting points vary, but seemingly not to accommodate for the length of the upcoming prosodic phrase. An explanation to the variation in F0 starting points was sought in the discourse. More specifically, we investigated whether the downward scaling of F0 register over sequences of several prosodic phrases is used to signal coherence among phrases in spontaneous speech. Moving from investigations of the tonal signaling of coherence

within the prosodic phrase, we thereby turned to tonal coherence signaling among prosodic phrases.

It was concluded that intonation is also used to signal coherence among prosodic phrases in spontaneous speech. Groups of prosodic phrases with uninterrupted downward scaling of F0 register and perceptually weak internal boundaries were found, although it was also observed that a downward scaling of F0 register in successive prosodic phrases (as observed in the difference in height between phrase-initial accent peaks and/or F0 start values) did not invariably give rise to perceptually weak phrase boundaries. It was consequently suggested that the presence of other cues also affect the degree of perceived boundary strength. In addition to other considerations, the observation that sequences of prosodic phrases separated by perceptually weak boundaries and delimited by strong boundaries, i.e. per definition 'prosodic utterances', far from always demonstrated an internal tonal coherence led us to analyze tonal coherence signaling across phrase boundaries as controlled by discourse structure. The downward scaling of F0 register observed in sequences of prosodic phrases that clearly had not been entirely planned in advance, or that had been reorganized during production, was claimed to give further support to the conclusion drawn in chapter four, namely that the downward trend of F0 need not be seen as requiring hard preplanning.

In chapter six, we followed up on the finding that intonation alone cannot explain the perceptual impression of coherence among prosodic phrases with two perception experiments. The purpose of the experiments was, firstly, to determine what the relationship is between perceived boundary strength and three known correlates of prosodic phrasing in Swedish. Although our spontaneous speech material allowed us to make only rather crude acoustic measurements, a vast effect of pause length on perceived boundary strength was shown to exist in spontaneous speech. Furthermore, extent of F0 reset appeared to have some effect upon perceived boundary strength whereas amount of final lengthening (change in articulation rate phrase-finally) seemed to matter little for the percept of boundary strength. Syntactic boundary type was another factor that was found to have a small effect on perceived boundary strength.

Secondly, we searched for support of a division into two phrasal categories or constituents in spontaneous Swedish. After a discussion based on findings from previous studies, we concluded that the intermediate phrase is not a relevant phrasal constituent in the description of Swedish, and moved on to examine the distinction made between the 'prosodic phrase' and the 'prosodic utterance' in the Swedish

intonation model. The results of the perception experiments showed that listeners' perception of boundary strength in spontaneous speech is continuous rather than categorical. In the evaluation of the base prosody system, it was proposed that the definitions of the different prosodic categories be refined, since it is felt that the existing system to some extent is unclear. We argued that the vagueness of the definitions when applied to spontaneous speech data is related to the continuous nature of boundary strength. As there is no separate edge tone or tonal head to distinguish the prosodic utterance from the prosodic phrase, the definitions of the 'prosodic phrase' and the 'prosodic utterance' rest largely on listeners' ability to make a categorical distinction between weak and strong boundaries. The listeners' high agreement on perceived boundary strength on a continuous scale indicates that quite many degrees of boundary strength can be produced and perceived in spontaneous speech. Based on this finding, we proposed that degree of boundary strength at this level of structure be described in the same manner as emphasis in the Swedish intonation model, namely as variation within one and the same phonological category. It was found that a communicatively more relevant feature of the prosodic boundary than its strength (as mainly signaled by pause length) in spontaneous speech was the degree of tonal coherence or continuity perceived across it. Whereas these two features of prosodic boundaries – degree of boundary strength and perceived coherence across the boundary – co-vary in read speech and motivate a categorical distinction between two phrasal constituents (definable in terms of boundary strength), the situation in spontaneous speech is clearly more complex. Future studies of spontaneous speech will most likely reveal further differences like this between read and spontaneous speech.

# References

Aasa, A., G. Bruce, O. Engstrand, A. Eriksson, M. Segerup, E. Strangert, I. Thelander and P. Wretling. (2000). Collecting dialect data and making use of them: an interim report from Swedia 2000. In: Botinis, A. and N. Torstensson (eds), *Fonetik 2000,* 17-20. Skövde: Department of Languages, University of Skövde.

Abney, S. A. (1987). *The English noun phrase in its sentential aspect*. Doctoral dissertation. MIT. Cambridge, Massachusetts.

Alstermark, M. and Y. Erikson. (1971). Swedish word accents as a function of word length. *Speech Transmission Laboratory Quarterly Status Report* 1/1971, 1-13. Stockholm: Department of Speech Communication and Speech Transmission Laboratory, Royal Institute of Technology.

Armstrong, L. and I. Ward. (1926). *A handbook of English intonation*. Leipzig and Berlin: Teubner.

Ashby, M. (1978). A study of two English nuclear tones. *Language and Speech* 21, 326-336.

Bannert, R. (1982). An F0-dependent model for duration? *RUUL* 8, 58-80. Uppsala: Department of Linguistics, University of Uppsala.

Bannert, R. and A.-C. Bredvad-Jensen. (1975). Temporal organization of Swedish tonal accents: The effect of vowel duration. *Working Papers* 10, 1-16. Lund: Department of Linguistics and Phonetics, Lund University.

Beckman, M. E. (1993). Modeling the Production of Prosody. In: House, D. and P. Touati (eds), *Proceedings of an ESCA Workshop on Prosody*. Working Papers 41, 258-263. Lund: Department of Linguistics and Phonetics, Lund University.

Beckman, M. E. (1997). A Typology of Spontaneous Speech. In: Sagisaka, Y., N. Campbell and N. Higuchi (eds), *Computing Prosody. Computational Models for Processing Spontaneous Speech*, 7-26. New York: Springer.

Beckman, M. E. and G. M. Ayers. (1993). *Guidelines for ToBI Labelling. Version 3*. The Ohio State University Research Foundation. Department of Linguistics, Ohio State University.

Beckman, M. E. and J. B. Pierrehumbert. (1985). Synthesizing Japanese using a downstep model. *The Journal of the Acoustical Society of America* 77, S 38.

Beckman, M. E. and J. B. Pierrehumbert. (1986). Intonational structure in Japanese and English. In: Ewin, C. and J. Anderson (eds), *Phonology Yearbook* 3, 255-309. Cambridge: Cambridge University Press.

Bing, J. (1985). *Aspects of prosody*. New York: Garland.

Bolinger, D. (1962). Intonation as a universal. *Proceedings of the Ninth International Congress of Linguistics*, 833-848. The Hague: Mouton & Co.

Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language* 48, 633-644.

Bruce, G. (1977). *Swedish Word Accent in Sentence Perspective*. Travaux de l'Institute de Linguistique de Lund 11. Doctoral dissertation. Lund: CWK Gleerup.

Bruce, G. (1981). Tonal and temporal interplay. In: Fretheim, T. (ed), *Nordic Prosody II: Papers from a symposium*, 63-74. Trondheim: Tapir.

Bruce, G. (1982a). Developing the Swedish Intonation Model. *Working Papers* 22, 51-116. Lund: Department of Linguistics and Phonetics, Lund University.

Bruce, G. (1982b). Textual Aspects of Prosody in Swedish. *Phonetica* 39, 274-287.

Bruce, G. (1984). Aspects of F0 Declination in Swedish. *Working Papers* 27, 51-64. Lund: Department of Linguistics and Phonetics, Lund University.

Bruce, G. (1987). How floating is focal accent? In: Gregersen, K. and H. Basbøll (eds), *Nordic Prosody IV: Papers from a symposium,* 41-49. Odense: Odense University Press.

Bruce, G. (1994). Prosodisk strukturering i dialog. In: Holmberg, A. and K. Larsson (eds), *Svenskans beskrivning* 20, 9-23. Lund: Lund University Press.

Bruce, G. (1998). *Allmän och svensk prosodi.* Praktisk lingvistik 16. Lund: Department of Linguistics and Phonetics, Lund University.

Bruce, G. (2001). Secondary stress and pitch accent synchronization in Swedish. In: van Dommelen, W. A. and T. Fretheim (eds), *Nordic Prosody: Proceedings of the VIIth conference*, 33-44. Frankfurt am Main: Peter Lang.

Bruce, G., C.-C. Elert, O. Engstrand, A. Eriksson and P. Wretling. (1999). Database tools for a prosodic analysis of the Swedish dialects. In: Andersson, R., Å. Abelin, J. Allwood and P. Lindblad (eds), *Fonetik 99*. Gothenburg Papers in Theoretical Linguistics 81, 37-40. Department of Linguistics, Göteborg University.

Bruce, G. and E. Gårding. (1978). A Prosodic Typology for Swedish Dialects. In: Gårding, E., G. Bruce and R. Bannert (eds), *Nordic Prosody: Papers from a symposium*, 219-228. Malmö: Department of Linguistics and Phonetics, Lund University.

Bruce, G. and B. Granström. (1993). Prosodic modelling in Swedish speech synthesis. *Speech Communication* 13, 63-73.

Bruce, G., B. Granström, K. Gustafson and D. House. (1991). Prosodic Phrasing in Swedish. *Working Papers* 38, 5-17. Lund: Department of Linguistics and Phonetics, Lund University.

Bruce, G., B. Granström, K. Gustafson and D. House. (1993). Interaction of F0 and duration in the perception of prosodic phrasing in Swedish. In: Granström, B. and L. Nord (eds), *Nordic Prosody VI: Papers from a symposium*, 7-22. Stockholm: Almqvist & Wiksell International.

Bruce, G., B. Granström, K. Gustafson, D. House and P. Touati. (1994). Preliminary Report from the Project "Prosodic Segmentation and Structuring of Dialogue". In: Bruce, G., D. House and P. Touati. (eds), *Fonetik –94.* Working Papers 43, 34-37. Lund: Department of Linguistics and Phonetics, Lund University.

Chomsky, N. and M. Halle. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Clark, H. H. (1994). Managing problems in speaking. *Speech Communication* 15, 243-250.

Collier, R. (1975). Physiological correlates of intonation patterns. *The Journal of the Acoustical Society of America* 58, 249-255.

Collier, R., A. Cohen and J. 't Hart. (1982). Declination: Construct or Intrinsic Feature of Speech Pitch? *Phonetica* 39, 254-273.

Collier, R. and J. 't Hart. (1975). The role of intonation in speech perception. In: Cohen, A. and S. G. Nooteboom (eds), *Structure and process in speech perception*, 107-121. Heidelberg: Springer-Verlag.

Cooper, W. E. (1976). *Syntactic Control of Speech Timing.* Doctoral dissertation. Cambridge, Massachusetts: MIT Press.

Cooper, W. E. and J. M. Sorensen. (1981). *Fundamental Frequency in Sentence Production.* New York: Springer-Verlag.

Cruttenden, A. (1986). *Intonation*. Cambridge : Cambridge University Press.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

Cutler, A., D. Dahan and W. van Donselaar. (1997). Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Speech* 40, 141-201.

Dahl, Ö. (1976). What is new information? In: Enkvist, N. E. and V. Kohonen (eds), *Reports on Text Linguistics: Approaches to Word Order*. Publications of the Research Institute of the Åbo Akademi Foundation 8, 37-50. Åbo.

Dankovičová, J. (1997). The domain of articulation rate variation in Czech. *Journal of Phonetics* 25, 287-312.

Delais-Roussarie, E. (1996). Phonological Phrasing and Accentuation in French. In: Nespor, M. and N. Smith (eds), *Dam Phonology: HIL Phonology Paper II*, 1-38. The Hague: Holland Academic Graphics.

Fant, G. and A. Kruckenberg. (1994). Notes on Stress and Word Accents in Swedish. *Speech Transmission Laboratory Quarterly Status Report* 2-3/1994, 125-144. Stockholm: Department of Speech Communication and Music Acoustics, Royal Institute of Technology.

Fletcher, J., E. Grabe and P. Warren. Forthcoming. Intonational variation in four dialects of English: the high rising tune. In: Jun, S.-A. (ed), *Prosodic Typology: Through Intonational Phonology and Transcriptions*. Oxford University Press.

Fox Tree, J. E. and H. H. Clark. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition* 62, 151-167.

Fretheim, T. (1981). Intonational Phrasing in Norwegian. *Nordic Journal of Linguistics* 4, 111-137.

Fretheim, T. (1991). Intonational phrasing and syntactic focus domains. In: Verschuerenm, J. (ed), *Levels of linguistic adaption*, 81-111. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Fretheim, T. (2001). The interaction of right-dislocated pronominals and intonational phrasing in Norwegian. In: van Dommelen, W. A. and T. Fretheim (eds), *Nordic Prosody: Proceedings of the VIIIth conference*, 61-75. Frankfurt am Main: Peter Lang.

Gårding, E. (1967a). *Internal Juncture in Swedish*. Travaux de l'Institute de Linguistique de Lund 4. Doctoral dissertation. Lund: CWK Gleerup.

Gårding, E. (1967b). Prosodiska drag i spontant och uppläst tal. In: Holm, G. (ed), *Svenskt talspråk*, 40-85. Stockholm: Almqwist & Wiksell.

Gårding, E. (1974). Den efterhängsna prosodin. In: Teleman, U. and T. G. Hultman (eds), *Språket i bruk*, 50-71. Lund: Gleerups.

Gårding, E. (1983). A Generative Model of Intonation. In: Cutler, A. and D. R. Ladd (eds), *Prosody: Models and Measurements*, 11-25. Germany: Springer-Verlag.

Gårding, E. (1987). How many intonation models are there in Lund? *Working Papers* 31, 1-10. Lund: Department of Linguistics and Phonetics, Lund University.

Gårding, E., R. Bannert, A.-C. Bredvad-Jensen, G. Bruce and K. Nauclér. (1974). Talar skåningarna svenska? In: Platzack, C. (ed), *Svenskans beskrivning* 8, 107-123. Lund: Department of Scandinavian languages, Lund University.

Gårding, E. and G. Bruce. (1981). A presentation of the Lund model for Swedish intonation. *Working Papers* 21, 69-76. Lund: Department of Linguistics and Phonetics, Lund University.

Gårding, E. and A. Eriksson. (1989). Perceptual cues to some Swedish prosodic phrase patterns – A peak shift experiment. *Speech Transmission Laboratory Quarterly Status Report 1/1989*, 13-16. Stockholm: Department of Speech Communication and Music Acoustics, Royal Institute of Technology.

Gårding, E. and D. House. (1985). Frasintonation, särskilt i svenska. In: Allén, S., L.-G. Andersson, J. Löfström, K. Nordenstam and B. Ralph (eds), *Svenskans beskrivning* 15, 205-221. Göteborg: Göteborgs universitet.

Gårding, E. and D. House. (1986). Production and perception of phrases in some Nordic dialects. *Working Papers* 29, 91- 114. Lund: Department of Linguistics and Phonetics, Lund University.

Gårding, E. and P. Lindblad. (1973). Constancy and variation in Swedish word accent patterns. *Working Papers* 7, 36-110. Lund: Department of Linguistics and Phonetics, Lund University.

Goldman-Eisler, F. (1972). Pauses, Clauses, Sentences. *Language and Speech* 15, 103-113.

Grabe, E. (1998). *Comparative Intonational Phonology: English and German*. MPI Series in Psycholinguistics 7. Doctoral dissertation. Wageningen, Ponsen en Looien.

Grabe, E., B. Post and F. Nolan. (2001). Modelling intonational Variation in English. The IViE system. In: Puppel, S. and G. Demenko (eds), *Proceedings of Prosody 2000,* 51-57. Poznan, Poland: Adam Mickiewitz University.

Granqvist, S. (1996). Enhancements to the Visual Analogue Scale, VAS, for listening tests. *TMH-QPSR*, 4/1996, 61-62. Stockholm: Department of Speech, Music and Hearing, Royal Institute of Technology.

Grice, M., D. R. Ladd and A. Arvaniti. (2000). On the place of phrase accents in intonational phonology. *Phonology* 17, 143-185.

Grice, M., M. Reyelt, R. Benzmüller, J. Mayer and A. Batliner. (1996). Consistency in transcription and labelling of German intonation with GToBI. In: Bunnell, H. T. and W. Idsardi (eds), *Proceedings ICSLP 96*, 1716-1719. Philadelphia, Pennsylvania: University of Delaware and Alfred I. duPont Institute.

Grønnum Thorsen, N. (1988). Intonation on Bornholm – Between Danish and Swedish. *Annual Report of the Institute of Phonetics, University of Copenhagen,* 25-138.

Grosjean, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics* 21, 501-529.

Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics* 19, 377-417.

Gussenhoven, C. (2002). Intonation and Interpretation: Phonetics and Phonology. In: Bernard, B. and I. Marlien (eds), *Speech Prosody 2002*, 47-57. Aix-en-Provence: Laboratoire Parole et Langage, Université de Provence.

Gussenhoven, C. and G. Bruce. (1999). Word prosody and intonation. In: van der Hulst, H. (ed), *Word Prosodic Systems in the Languages of Europe*, 233-271. Berlin: Mouton de Gruyter.

Gussenhoven, C. and A. C. M. Rietveld. (1988). Fundamental frequency declination in Dutch: testing three hypotheses. *Journal of Phonetics* 16, 355-369.

Gussenhoven, C. and A. C. M. Rietveld. (1992). Intonation contours, prosodic structure and preboundary lengthening. *Journal of Phonetics* 20, 283-303.

Hadding-Koch, K. (1961). *Acoustico-Phonetic Studies in the Intonation of Southern Swedish.* Travaux de l'Institute de Linguistique de Lund 3. Doctoral dissertation. Lund: CWK Gleerup.

Haegeman, L. (1994). *Introduction to government & binding theory*. Second edition. Cambridge, Massachusetts: Blackwell.

Halliday, M. A. K. (1967). Notes on Transitivity and Theme in English 2. *Journal of Linguistics* 3, 199-244.

Hansson, P. (2001). The effect of individual words' information status on accentuation. In: van Dommelen, W. A. and T. Fretheim (eds), *Nordic Prosody: Proceedings of the VIIIth conference*, 89-101. Frankfurt am Main: Peter Lang.

Hansson, P. (2002). Articulation Rate Variation in South Swedish Phrases. In: Bernard, B. and I. Marlien (eds), *Speech Prosody 2002*, 371-374. Aix-en-Provence: Laboratoire Parole et Langage, Université de Provence.

Harris, M. O., N. Umeda and J. Bourne. (1981). Boundary perception in fluent speech. *Journal of Phonetics* 9, 1-18.

Hart, J. 't. (1986). Declination has not been defeated – A reply to Lieberman et al. *The Journal of the Acoustical Society of America* 80, 1838-1840.

Heeman, P. A. (1997). *Speech Repairs, Intonational Boundaries and Discourse Markers: Modeling Speaker's Utterances in Spoken Dialogue*. Technical Report 673. Doctoral dissertation. Computer Science Department, University of Rochester.

Heldner, M. (2001). *Focal accent – $f_0$ movements and beyond.* PHONUM 8. Reports in Phonetics Umeå University. Doctoral dissertation. Umeå: Department of Phonetics, Umeå University.

Hirschberg, J. (1999). Communication and prosody: Functional aspects of prosody. *Proceedings of the ESCA workshop on Dialogue and Prosody*, 7-15. Veldhoven, The Netherlands.

Hirschberg, J. and J. B. Pierrehumbert. (1986). Intonational Structuring of Discourse. *Proceedings of the 24th Meeting of the Association for Computational Linguistics*, 136-144. New York.

Horne, M. (1994). Generating prosodic structure for synthesis of Swedish intonation. In: Bruce, G., D. House and P. Touati (eds), *Fonetik –94.* Working Papers 43, 72-75. Lund: Department of Linguistics and Phonetics, Lund University.

Horne, M. and M. Filipsson. (1998). From prosodic structure to intonation contours. In: Werner, S. (ed), *Nordic Prosody: Proceedings of the VIIth conference,* 127-139. Framkfurt am Main: Peter Lang.

Horne, M., P. Hansson, G. Bruce and J. Frid. (2001). Accent Patterning on Domain-Related Information in Swedish Travel Dialogues. *International Journal of Speech Technology* 4, 93-102.

Horne, M., P. Hansson, G. Bruce, J. Frid and M. Filipsson. (2001). Cue words and the topic structure of spoken discourse: The case of Swedish men 'but'. *Journal of Pragmatics* 33, 1061-1081.

Horne, M., E. Strangert and M. Heldner. (1995). Prosodic Boundary Strength in Swedish: Final Lengthening and Silent Interval Duration. In: Elenius, K. and P. Branderud (eds), *ICPhS 95,* 170-173. Stockholm: Department of Speech Communication and Music Acoustics, Royal Institute of Technology and Department of Linguistics, Stockholm University.

House, D. (1985). Sentence prosody and syntax in speech perception. *Working Papers* 28, 91-108. Lund: Department of Linguistics and Phonetics, Lund University.

House, D. (1990). *Tonal Perception in Speech*. Travaux de l'Institute de Linguistique de Lund 24. Doctoral dissertation. Lund: Lund University Press.

House, A. and G. Fairbanks. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America* 25, 105-113.

House, D., D. Hermes and F. Beaugendre. (1997). Temporal-alignment categories of accent-lending rises and falls. In: Kokkinakis, G., N. Fakotakis and E. Dermatas (eds), *Eurospeech '97 Proceedings*, 879-882. Rhodes, Greece: Wire Communications Laboratory, University of Patras.

House, D., D. Hermes and F. Beaugendre. (1998). Perception of tonal rises and falls for accentuation and phrasing in Swedish. In: Mannell, R. H. and J. Robert-Ribes (eds), *ICSLP '98 Proceedings*, 2799-2802. Sydney, Australia: Australian Speech Science and Technology Association.

Huber, D. (1988). *Aspects of the communicative function of voice in text intonation. Constancy and variability in Swedish fundamental frequency contours*. Doctoral dissertation. Göteborg: Department of Computational Linguistics, Göteborg University.

Jespersen, O. (1924). *The philosophy of grammar*. London: Allen & Unwin.

Jun, S.-A. (2002). Syntax over focus. In: *ICSLP 2002*, 2281-2284. Denver, Colorado: Center for Spoken Language Research and Department of Speech,

Language, Hearing Sciences, University of Colorado and W.J. Gould Voice Center, Denver Center for the Performing Arts.

Kang, S. and S. Speer. (2002). Prosody and Clause Boundaries in Korean. In: Bernard, B. and I. Marlien (eds), *Speech Prosody 2002*, 419-421. Aix-en-Provence: Laboratoire Parole et Langage, Université de Provence.

Klatt, D. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics* 3, 129-149.

Ladd, D. R. (1983). Peak Features and Overall Slope. In: Cutler, A. and D. R. Ladd (eds), *Prosody: Models and Measurements*, 39-52. Germany: Springer-Verlag.

Ladd, D. R. (1986). Intonational phrasing: the case for recursive prosodic structure. In: Ewin, C. and J. Anderson (eds), *Phonology Yearbook* 3, 311-340. Cambridge: Cambridge University Press.

Ladd, D. R. (1996) *Intonational Phonology. Cambridge Studies in Linguistics 79*, Cambridge: Cambridge University Press.

Ladd, D. R. and R. Morton. (1997). The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics* 25, 313-342.

Lambrecht, K. (1994). *Information structure and sentence form. Topic, focus, and the mental representation of discourse referents.* Cambridge Studies in Linguistics 71. Cambridge: Cambridge University Press.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Massachusetts: MIT Press.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa* 7, 107-122.

Lehiste, I. (1975). The phonetic structure of paragraphs. In: Cohen, A. and S. G. Nooteboom (eds), *Structure and Process in speech perception*, 165-203. Cambridge, Massachusetts: Springer.

Lehiste, I and G. E. Peterson. (1961). Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America* 31, 428-435.

Liberman, M. (1975). *The intonational system of English*. Doctoral dissertation. MIT. Distributed 1978 by the Indiana University Linguistics Club.

Liberman, M and J. B. Pierrehumbert. (1984). Intonational Invariance under Changes in Pitch Range and Length. In: Aronoff, M. and R. T. Oehrle (eds), *Studies in Phonology Presented to Morris Halle by His Teacher and Students*, 157-233. Cambridge, Massachusetts: MIT Press.

Liberman, M. and A. Prince. (1977). On stress and linguistic rhythm. *Linguistic Inquiry* 8, 249-336.

Lieberman, P. (1967). *Intonation, Perception and Language*. Cambridge, Massachusetts: MIT Press.

Lieberman, P. (1986). Alice in declinationland – A reply to Johan 't Hart. *The Journal of the Acoustical Society of America* 80, 1840-1842.

Lieberman, P., W. Katz, A. Jongman, R. Zimmerman and M. Miller. (1985). Measures of the sentence intonation of read and spontaneous speech in American English. *The Journal of the Acoustical Society of America* 77, 649-657.

Lindblom, B. (1978). Final lengthening in speech and music. In: Gårding, E., G. Bruce and R. Bannert (eds), *Nordic Prosody: Papers from a symposium*, 85-101. Malmö: Department of Linguistics and Phonetics, Lund University.

Lindblom, B., B. Lyberg and K. Holmgren. (1976). *Durational Patterns of Swedish Phonology: Do They Reflect Short-Term Motor Memory Processes?* Published in: Lyberg, Bertil. (1981). Temporal properties of spoken Swedish. Monographs from the Institute of Linguistics, University of Stockholm 6. Doctoral dissertation.

Lindström, A., I. Bretan and M. Ljungqvist. (1996). Prosody Generation in Text-to-Speech Conversion Using Dependency Graphs. In: Bunnell, H. T. and W. Idsardi (eds), *Proceedings ICSLP 96*, 1341-1344. Philadelphia, Pennsylvania: University of Delaware and Alfred I. duPont Institute.

Linell, P. (in preparation). *En dialogisk grammatik?* Unpublished manuscript. The Tema Institute (Tema Kommunikation), Linköping University.

Löfqvist, A. (1975). Intrinsic and Extrinsic F0 variations in Swedish Tonal Accents. *Phonetica* 31, 228-247.

Lyberg, B. (1977). Some observations on the timing of Swedish utterances. *Journal of Phonetics* 5, 49-59.

Lyberg, B. (1978). Final lengthening – A consequence of articulatory and perceptual restrictions? *Annual Report of the Institute of Phonetics, University of Copenhagen*, 79-85.

Lyberg, B. (1981). Some consequences of a model for segment duration based on F0-dependence. *Journal of Phonetics* 9, 97-103.

Lyberg, B. (1984). Some fundamental frequency perturbations in a sentence context. *Journal of Phonetics* 12, 307-317.

Maeda, S. (1976). *A characterization of American English intonation*. Doctoral dissertation. MIT. Cambridge, Massachusetts.

Malmberg, B. (1963). *Structural linguistics and human communication*. Berlin: Springer.

Mayo, C., M. Aylett and D. R. Ladd. (1997). Prosodic transcription of Glasgow English: An evaluation study of GlaToBI. In: Botonis, A., G. Kouroupetroglou and G. Carayiannis (eds), *Proceedings of an ESCA workshop: Intonation: Theory, Models and Applications*, 231-234. Athens, Greece: ESCA and the University of Athens.

McCarthy, J. J. and A. Prince. (1993). Generalized alignment. In: Booij, G. and J. van Marle (eds.), *Yearbook of Morphology 1993,* 79-153. Dordrecht: Kluwer Academic Publishers.

Molnár, V. (1998). Topic in focus. On the syntax, phonology, semantics and pragmatics of the so-called "contrastive topic" in Hungarian and German. *Acta Linguistica Hungarica* 45, 89-166.

Nakatani, C. H. and J. Hirschberg. (1994). A corpus-based study of repair cues in spontaneous speech. *The Journal of the Acoustical Society of America* 95, 1603-1616.

Nespor, M. and I. Vogel. (1986). *Prosodic Phonology*. Dordrecht: Foris Publications.

Nolan, F. (1995). The effect of emphasis on declination in English intonation. In: Windsor Lewis, J. (ed), *Studies in General and English Phonetics. Essays in Honour of Professor J. D. O'Connor*, 241-254. London: Routledge.

Ohala, J. J. and B. W. Eukel. (1987). Explaining the Intrinsic Pitch of Vowels. In: Channon, R. and L. Shockey (eds), *In Honor of Ilse Lehiste/Ilse Lehiste Puhendusteos*, 207-215. Dordrecht: Foris.

Öhman, S. (1967). Word and sentence intonation: a quantitative model. *Speech Transmission Laboratory Quarterly Status Report* 2/1967, 20-54. Stockholm: Department of Speech Communication and Music Acoustics, Royal Institute of Technology.

O'Shaughnessy, D. and J. Allen. (1983). Linguistic modality effects on fundamental frequency in speech. *The Journal of the Acoustical Society of America* 74, 1155-1171.

Pierrehumbert, J. B. (1979). The perception of fundamental frequency declination. *The Journal of the Acoustical Society of America* 66, 363-369.

Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation*. Doctoral dissertation. MIT. Cambridge, Massachusetts. Reproduced and distributed 1987 by the Indiana University Linguistics Club. Bloomington, Indiana.

Pierrehumbert, J. B. and M. E. Beckman. (1988). *Japanese Tone Structure.* Linguistics inquiry monographs 15. Cambridge, Massachusetts: MIT Press.

Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Pitrelli, J. F., M. E. Beckman and J. Hirschberg. (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. In: *ICSLP 94*, 123-126. Yokohama, Japan: The Acoustical Society of Japan.

Prieto, P. (1998). The scaling of the L values in Spanish downstepping contours. *Journal of Phonetics* 26, 261-282.

Prieto, P., C. Shih and H. Nibert. (1996). Pitch downtrend in Spanish. *Journal of Phonetics* 24, 445-473.

Riad, T. (1998). Towards a Scandinavian accent typology. In: Kehrein, W. and R. Wiese (eds), *Phonology and Morphology of the Germanic Languages*. Linguistische Arbeiten 386, 77-109. Tübingen: Niemeyer.

Rietveld, A. C. M. and C. Gussenhoven. (1985). On the relation between pitch excursion size and pitch prominence. *Journal of Phonetics* 13, 299-308.

Rietveld, A. C. M. and C. Gussenhoven. (1987). Perceived speech rate and intonation. *Journal of Phonetics* 15, 273-285.

Sanderman, A. A. (1996). *Prosodic phrasing. Production, perception, acceptability and comprehension.* Doctoral dissertation. Eindhoven: Institute for Perception Research, Eindhoven University of Technology.

Sanderman, A. A. and R. Collier. (1997). Prosodic phrasing and comprehension. *Language and Speech* 40, 391-409.

Schriberg, E. (1994). *Preliminaries to a Theory of Speech Disfluencies*. Doctoral dissertation. University of California, Berkeley.

Selkirk, E. (1984). *Phonology and syntax: the relation between sound and structure.* Cambridge, Massachusetts: MIT Press.

Selkirk, E. (2000). The Interaction of Constraints on Prosodic Phrasing. In: Horne, M. (ed), *Prosody. Theory and Experiment. Studies presented to Gösta Bruce*, 231-261. Dordrecht: Kluwer Academic Publishers.

Shattuck-Hufnagel, S. and A. E. Turk. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25, 193-247.

Silverman, K., M. E. Beckman, J. Pitrelli, M. Ostendorf, C. W. Wightman, P. Price, J. B. Pierrehumbert and J. Hirschberg. (1992). ToBI: a standard for labeling English prosody. In: Ohala, J. J., T. M. Nearey, B. L. Derwing, M. M. Hodge and G. E. Wiebe (eds), *ICSLP 92 Proceedings*, 867-70. Banff, Canada: University of Alberta.

Strangert, E. (1993). Speaking style and pausing. In: Strangert, E., M. Heldner and P. Czigler (eds), *PHONUM* 2, 121-137. Umeå: Department of Phonetics, University of Umeå.

Strangert, E. and M. Heldner. (1995a). Labelling of boundaries and prominences by phonetically experienced and non-experienced transcribers. *PHONUM* 3, 85-109. Umeå: Department of Phonetics, Umeå University.

Strangert, E. and M. Heldner. (1995b). The labeling of prominence in Swedish by phonetically experienced transcribers. In: Elenius, K. and P. Branderud (eds), *ICPhS 95,* 204-207. Stockholm: Department of Speech Communication and Music Acoustics, Royal Institute of Technology and Department of Linguistics, Stockholm University.

Strawson, P. (1964). Identifying reference and truth values. *Theoria* 30, 96-118.

Stubbs, M. (1983). *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*. Oxford: Basil Blackwell.

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30, 485-496.

Swerts, M., E. Strangert and M. Heldner. (1996). F0 declination in read-aloud and spontaneous speech. In: Bunnell, H. T. and W. Idsardi (eds), *Proceedings ICSLP 96*, 1501-1504. Philadelphia, Pennsylvania: University of Delaware and Alfred I. duPont Institute.

Tench, P. (1995). The boundaries of intonation units. In: Windsor Lewis, J. (ed), *Studies in General and English Phonetics. Essays in Honour of Professor J.D. O'Conner*, 270-277. London: Routledge.

Terken, J. M. B. and J. Hirschberg. (1994). Deaccentuation of Words Representing 'Given' Information: Effects of persistence of Grammatical Function and Surface Position. *Language and Speech* 37, 125-145.

Thorsen, N. (1980). A study of the perception of sentence intonation – evidence from Danish. *The Journal of the Acoustical Society of America* 67, 1014-1030.

Thorsen, N. (1983). Two issues in the Prosody of Standard Danish. In: Cutler, A. and D. R. Ladd (eds), *Prosody: Models and Measurements*, 27-38. Germany: Springer-Verlag.

Touati, P. (1988). *Structures prosodiques du suédois et du français. Profiles temporels et configurations tonales.* Travaux de l'Institute de Linguistique de Lund 21. Doctoral dissertation. Lund: Lund University Press.

Truckenbrodt, H. (1995). *Phonological phrases: Their relation to syntax, focus, and prominence*. Doctoral dissertation. MIT. Cambridge, Massachusetts.

Truckenbrodt, H. (1999). On the Relation between Syntactic Phrases and Phonological Phrases. *Linguistic Inquiry* 30, 219-255.

Umeda, N. and A. M. S. Quinn. (1981). Word duration as an acoustic measure of boundary perception. *Journal of Phonetics* 9, 19-28.

van den Berg, R., C. Gussenhoven and A. C. M. Rietveld. (1992). Downstep in Dutch: implications for a model. In: Docherty, G. J. and D. R. Ladd (eds), *Papers in Laboratory Phonology. Gesture, Segment, Prosody*, 335-359. Cambridge: Cambridge University Press.

Vanderslice, R. and P. Ladefoged. (1972). Binary suprasegmental features and transformational word-accentuation rules. *Language* 48, 819-839.

Wewers, M. E. and N. K. Lowe. (1990). A Critical Review of Visual Analogue Scales in the Measurement of Clinical Phenomena. *Research in Nursing & Health* 13, 227-236.

Whalen, D. H. and A. G. Levitt. (1995). The Universality of Intrinsic F0 of Vowels. *Journal of Phonetics* 23, 349-366.

Wightman, C. W. (2002). ToBI Or Not ToBI? In: Bernard, B. and I. Marlien (eds), *Speech Prosody 2002*, 25-30. Aix-en-Provence: Laboratoire Parole et Langage, Université de Provence.

Wightman, C. W., S. Shattuck-Hufnagel, M. Ostendorf and P. J. Price (1992). Segmental Durations in the Vicinity of Prosodic Phrase Boundaries. *The Journal of the Acoustical Society of America* 91, 1707-1717.

Zetterlund, S., L. Nordstrand and O. Engstrand. (1978). An experiment on the perceptual evaluation of prosodic parameters for phrase structure decision in Swedish. In: Gårding, E., G. Bruce and R. Bannert (eds), *Nordic Prosody: Papers from a symposium,* 15-32. Malmö: Department of Linguistics, Lund University.

*Web addresses*

Baumann, S. and M. Grice. (2002). *GToBI*. Computational Linguistics and Phonetics, Saarland University. http://www.coli.uni-sb.de/phonetik/ (links: 'Projects' and 'GToBI') (Accessed 2003-01-06).

Engstrand, O. (2000). *100 svenska dialekter idag!* Department of Linguistics and Phonetics, Lund University, Department of Linguistics, Stockholm University and Department of Phonetics, Umeå University. http//:www.swedia.nu (Accessed 2003-01-06).

Grabe, E. (2001). *The IViE Labelling Guide*. Version 3. Phonetics Laboratory, University of Oxford and Department of Linguistics, University of Cambridge. http://www.phon.ox.ac.uk/~esther/ivyweb/guide.html (Accessed 2003-01-06).

Gussenhoven, C., A. C. M. Rietveld and J. M. B. Terken. (1999). *ToDI. Transcription of Dutch Intonation.* Version 1.1 (First Edition. June 1999). ToDI Collective. http://lands.let.kun.nl/todi/todi/home.htm (Accessed 2003-01-06).

Jönsson, A. (1998). *Swedish Dialogue Systems.* http://www.ida.liu.se/~nlplab/sds/ (Accessed 2003-01-06).

The Ohio State University Department of Linguistics. (1999). *ToBI*. http://www.ling.ohio-state.edu/~tobi/ (Accessed 2003-01-06).

# TRAVAUX DE L'INSTITUT DE LINGUISTIQUE DE LUND

## FONDÉS PAR BERTIL MALMBERG.

1 *Carl-Gustaf Söderberg*. A Typological Study on the Phonetic Structure of English Words with an Instrumental-Phonetic Excursus on English Stress. 1959.
2 *Peter S. Green*. Consonant-Vowel Transitions. A Spectrographic Study. 1959.
3 *Kerstin Hadding-Koch*. Acoustico-Phonetic Studies in the Intonation of Southern Swedish. 1961.
4 *Börje Segerbäck*. La réalisation d'une opposition de tonèmes dans des dissyllabes chuchotés. Étude de phonétique expérimentale. 1966.
5 *Velta Ruke-Dravina*. Mehrsprachigkeit im Vorschulalter. 1967.
6 *Eva Gårding*. Internal Juncture in Swedish. 1967.
7 *Folke Strenger*. Les voyelles nasales françaises. 1969.
8 *Edward Carney*. Hiss Transitions and their Perception. 1970.
9 *Faith Ann Johansson*. Immigrant Swedish Phonology. 1973.
10 *Robert Bannert*. Mittelbairische Phonologie auf akustischer und perzeptorischer Grundlage. 1976.
11 *Eva Gårding*. The Scandinavian Word Accents. 1977.
12 *Gösta Bruce*. Swedish Word Accents in Sentence Perspective. 1977.
13 *Eva Gårding, Gösta Bruce, Robert Bannert* (eds.). Nordic Prosody. 1978.
14 *Ewa Söderpalm*. Speech Errors in Normal and Pathological Speech. 1979.
15 *Kerstin Nauclér*. Perspectives on Misspellings. 1980.
16 *Per Lindblad*. Svenskans sje- och tjeljud (Some Swedish sibilants). 1980.
17 *Eva Magnusson*. The Phonology of Language Disordered Children. 1983.
18 *Jan-Olof Svantesson*. Kammu Phonology and Morphology. 1983.
19 *Ulrika Nettelbladt*. Developmental Studies of Dysphonology in Children. 1983.
20 *Gisela Håkansson*. Teacher Talk. How Teachers Modify their Speech when Addressing Learners of Swedish as a Second Language. 1987.
21 *Paul Touati*. Structures prosodiques du suédois et du français. Profils temporels et configurations tonales. 1987.
22 *Antonis Botinis*. Stress and Prosodic Structure in Greek. A Phonological, Acoustic, Physiological and Perceptual Study. 1989.
23 *Karina Vamling*. Complementation in Georgian. 1989.
24 *David House*. Tonal Perception in Speech. 1990.
25 *Emilio Rivano Fischer*. Topology and Dynamics of Interactions - with Special Reference to Spanish and Mapudungu. 1991.
26 *Magnus Olsson*. Hungarian Phonology and Morphology. 1992.
27 *Yasuko Nagano-Madsen*. Mora and Prosodic Coordination. A Phonetic Study of Japanese, Eskimo and Yoruba. 1992.
28 *Barbara Gawronska*. An MT Oriented Model of Aspect and Article Semantics. 1993.
29 *Bengt Sigurd* (ed.). Computerized Grammars for Analysis and Machine Translation. 1994.
30 *Arthur Holmer*. A Parametric Grammar of Seediq. 1996.
31 *Ingmarie Mellenius*. The Acquisition of Nominal Compounding in Swedish. 1997.
32 *Christina Thornell.* The Sango Language and Its Lexicon (Sêndâ-yângâ tî Sängö). 1997.
33 *Duncan Markham.* Phonetic Imitation, Accent, and the Learner. 1997.
34 *Christer Johansson.* A View from Language. Growth of Language in Individuals and Populations. 1997.
35 *Marianne Gullberg.* Gesture as a Communication Strategy in Second Langauge Discourse. A Study of Learners of French and Swedish. 1998.
36 *Mechtild Tronnier*. Nasals and Nasalisation in Speech Production. With Special Emphasis on Methodology and Osaka Japanese. 1998.
37 *Ann Lindvall*. Transitivity in Discourse. A Comparison of Greek, Polish and Swedish. 1998.
38 *Kirsten Haastrup & Åke Viberg* (eds.). Perspectives on Lexical Acquisition in a Second Language. 1998.
39 *Arthur Holmer, Jan-Olof Svantesson & Åke Viberg* (eds). Proceedings of the 18th Scandinavian Conference of Linguistics. 2001.
40 *Caroline Willners*. Antonyms in Context. A Corpus-based Semantic Analysis of Swedish Descriptive Adjectives. 2001.
41 *Hong Gao*. The Physical Foundation of the Patterning of Physical Action Verbs. A Study of Chinese Verbs. 2001.
42 *Anna Flyman Mattsson*. Teaching, Learning, and Student Output. A Study of French in the Classroom. 2003.
43 *Petra Hansson*. Prosodic Phrasing in Spontaneous Swedish. 2003.