**Nuclear Intonation in Swedish : Evidence from Experimental-Phonetic Studies and a Comparison with German**

Ambrazaitis, Gilbert

2009

# Nuclear Intonation in Swedish

Evidence from Experimental-Phonetic Studies
and a Comparison with German

Gilbert Ambrazaitis

LUND
UNIVERSITY

# Acknowledgements

My two supervisors Gösta Bruce and Merle Horne have contributed to this thesis in many ways. Thank you, Gösta, for your insightful guidance and your enlightening perspectives and questions concerning my data ("But if you just look at the $F_0$ curve, can you see if it's a German or a Swedish one?"). Thanks to both of you, Gösta and Merle, for a lot of valuable advice, many fruitful discussions and helpful comments on my manuscripts.

I had the opportunity to work in an excellent research environment, equipped with some wonderful and supportive colleagues. Tack till hela korridoren 4 och 5! I am especially indebted to Victoria Johansson, Mikael Roll, Susanne Schötz, and Marcus Uneson for sharing their technical knowledge as well as many other helpful thoughts with me – not least in these last few days. I am also very grateful for all support I received from our administrative and technical staff, especially (in alphabetical order) Malgorzata Andréasson, Johan Dahl, Lukas Gödke, Birgitta Lastow, Stefan Lindgren, Birgitta Lundahl, and Britt Nordbeck. Tack!

Quite a lot of people had a hand in the studies presented in this thesis. At an initial stage of planning the materials, some friends and colleagues served as informants in numerous informal pilot tests. I believe the people who I bothered most were probably Katja Schlund, Mikael Roll, and Marcus Uneson. Mikael Roll also acted as the Swedish speaker of the context questions in the final material used in the elicitation studies. Förlåt att jag störde, och tack ska ni ha!

For two of my speech copora I used the BAS Speech Recorder. During the development of the recording script, I received much valuable advice from Christoph Draxler (Munich). I carried out most of my recordings at the Humanities Laboratory (Lund University). Susanne Schötz, Anders Sjöström, and Stefan Lindgren were very supportive in planning and building the recoding settings. I also made some recordings in Munich, where Eric Lukac managed the recording sessions, and Lasse Bombien recruited the speakers. Danke sehr, tack, to all of you!

The perception experiment reported in this thesis was mainly conducted in Stockholm, both at KTH and at Stockholm University. For providing me either with the necessary technical facilities, or assistance in finding a large number of listeners in a very short amount of time, I would like to express my gratitude to Hassan Djamshidpey, Rickard Franzén, Björn Granström, Lisa Gustavsson, and Sara Myrberg.

I also wish to thank Sofia Söderberg and, once more, Mikael Roll, who volunteered as 'stimulus speakers' in the perception experiment. From Joost van de Weijer I have received a lot of advice concerning the design and the statistical evaluation of my studies during the years. Tack!

Finally, my studies would not have been possible without all the students, colleagues, and friends who participated either as speakers in my recordings –

# Contents

# Chapter 1

# Introduction

"Swedish and Norwegian have considerably simpler intonation systems than West Germanic. Where English, Dutch, and German have a large number of pitch accents to signal various shades of information status (H*L, H*, L*H, etc. [...]), Stockholm Swedish has essentially only one intonation contour if we disregard the continuation rise, and so has Danish, while East Norwegian has just two, a declarative and an interrogative contour. Of course, the fact that Scandinavian has (had) lexical pitch accents can be related to the sparser inventory of intonational contrasts. However, the next chapter [on Central Franconian, ga] will show that intonation systems can be more elaborate than those of Scandinavian and still have a lexical tone contrast." (Gussenhoven 2004)

This thesis is concerned with *prosody*, more specifically with *intonation* and *word accents*. The primary target language is the Standard variety of Swedish as it is spoken in the Stockholm area (*Sveamål*), but Standard German, as spoken in Northern Germany, serves as a reference language. The point of departure is a problem that arises when the intonational system of Swedish is discussed in typological terms, i.e. when it is compared to the systems of other, especially the closely-related West Germanic languages, as discussed in 1.1.

## 1.1   Tone and intonation in Swedish

Languages use variations in pitch for different purposes. A basic distinction can be made between *lexical* (or morphological) usages of pitch, which are referred to by notions such as *tone* or *word accent*, and the usage of speech melody for pragmatic or syntactic, i.e. sentence- or discourse related purposes, which is referred to as *intonation*. While all languages are generally assumed to have *into-*

1

*nation*, about 60–70% of the world's languages (cf. Yip 2002) are estimated to have *tone*, in addition to intonation. Languages such as German or English, which use pitch *exclusively* for intonational purposes, are often referred to as *intonation languages*, while languages which exhibit at least some lexical usage of pitch can be referred to as *tone languages* (e.g. Yip 2002; Gussenhoven 2004).

However, tone languages, in this definition, vary greatly concerning the density of lexical tones and the number of possible tonal contrasts. Languages like Swedish, Norwegian, Serbo-Croatian, or Japanese, are sometimes treated as a third group of languages, or as a specific sub-type of tone languages, referred to as *pitch-accent* or *accentual* languages. They use pitch lexically or morphologically, but to a much lesser extent than the most prototypical tone languages like Mandarin, in which there is a choice between four tones, and basically every syllable is assigned one of them. In Swedish, instead, only the lexically stressed syllables are associated with a so-called *word accent*, and there is only a binary choice between *accent I* and *accent II*.

Although, as mentioned above, all languages of the world probably use pitch for sentence- or discourse-related purposes, the choice between different intonation patterns is usually assumed to be somewhat more restricted in tone languages as compared to intonation languages. More specifically, tone languages typically have a smaller inventory of *intonational pitch accents*[1] or even none at all (e.g. Yip 2002). However, tone languages typically still have a choice between different *boundary tones* (Yip 2002). Moreover, parameters such as pitch register or pitch range of the lexically-determined tones can be varied in different ways for intonational purposes, e.g. for signalling *focus* or interrogativity (e.g. Liu and Xu 2005 on Mandarin).

Also for Swedish, a relatively small repertoire of intonation patterns is typically assumed, as indicated by the quotation of Gussenhoven (2004) above. This assumed difference between Swedish on the one hand, and German or other West Germanic languages on the other hand, is reflected in a comparison of contemporary models of Swedish and German intonation. Table 1.1 displays one example each of phonological accounts of the two languages: the Lund model for Swedish (Bruce 1977; Bruce and Gårding 1978; Bruce et al. 1997; Bruce 2005), and GToBI for German (Grice et al. 2005). They have been chosen for this illustration because both are formulated in terms of Autosegmental-Metrical (AM) phonology (Goldsmith 1990; Ladd 1996) and should hence be formally comparable. According to Table 1.1, German has six different accents at the utterance level, while Swedish has only one, known as the *focal accent* (Bruce 2005) or originally the *sentence accent* (Bruce 1977). A similar relation holds for final boundary tones.

---

[1]The notions printed in italics in this paragraph are defined and discussed in more detail in later sections.

Table 1.1: The (essential elements of the) tonal inventories of two AM-models: GToBI for German (Grice et al. 2005) and the Lund Model for Swedish (Bruce et al. 1997; Bruce 2005).

| Function | Standard German | | Standard Swedish | |
|---|---|---|---|---|
| | *accents* | *final boundaries* | *accents* | *final boundaries* |
| *word level* | | | H+L* | |
| | | | H*+L | |
| *utterance level* | H* | L- | | |
| | L+H* | H- | H- | |
| | L* | L-% | | L% |
| | L*+H | L-H% | | LH% |
| | H+L* | H-% | | |
| | H+!H* | H-ˆH% | | |

However, the presence of the tonal word accents alone can hardly explain the assumed simplicity of the Swedish intonation system, since "intonation systems can be more elaborate than those of Scandinavian and still have a lexical tone contrast" (Gussenhoven 2004), which is demonstrated by Gussenhoven (2004) for the case of Central Franconian.

An alternative explanation for the observed difference between the intonation systems of German and Swedish, as assumed by the models in Table 1.1, is that these models are simply not equivalent, in the sense that they have been developed under different conditions and research paradigms. To anticipate Chapters 2 and 3, a larger variety of intonational *functions* have traditionally been taken into account in the modelling of German as compared to Swedish intonation. In particular, the Lund model for Swedish is intended to capture two basic prosodic functions: the signalling of *prominence* (exploited for the marking of *focus*) and *phrasing*. Attitudinal meanings, as also the signalling of "various shades of information status" (Gussenhoven 2004) have traditionally played a minor role in Swedish prosody research, while they have been treated to a great extent in the West Germanic traditions, as discussed in more detail in Chapter 3.

Hence, from a comparison of available models alone (e.g. Table 1.1), we cannot safely conclude that the intonation system of Swedish really differs from the systems of West Germanic languages. It might thus be the case that Swedish has an inventory of intonation patterns more similar to West Germanic languages than what is indicated by Table 1.1. In fact, listening informally to Swedish speech of-

ten provides the impression that there are more intonation patterns than the one or two mentioned by Gussenhoven (2004) and represented in the Lund model. For instance, even if questions may be typically spoken with a falling intonation in Swedish, similar to the one found in statements, rising intonation in questions can be observed, as well (as also suggested by Elert 2000).

In a recent paper, Féry (2009) discusses the typical classification of languages into intonation and tone languages as defined above and states that "[t]raditionally, languages have been divided on the basis of their word melodies". Féry (2009) suggests a new typology, which takes the inventory of sentence melodies, instead of word melodies, as a classification criterion. In her system, a language is classified as an *intonation language* if it exhibits a "rich array of pragmatically triggered phrasal tones", while it still may have some lexical usage of pitch. Hence, if indeed Swedish intonation were more similar to German intonation than typically assumed, then Féry's (2009) typology would seem to capture this similarity in an appropriate manner, since both Swedish and German would be classified as *intonation languages* on the basis of Féry's (2009) typology.

## 1.2  General goal, hypothesis, and scope of the thesis

The overall goal of the thesis is to contribute to a better understanding of Swedish intonation in the light of the problem sketched in 1.1. To this end, the intonational signalling of a number of selected pragmatic distinctions is compared in Swedish and German. The general hypothesis is that Swedish might have an inventory of intonation patterns similar to that of German, and that these patterns are also used for similar pragmatic purposes. Thus, the null hypothesis is that the intonation patterns that can be derived from the Lund model in fact suffice for an adequate phonological description of Swedish intonation.

Two important components of intonation are *accentuation* and *phrasing*, as already indicated by the mentioning of accents and boundary tones in 1.1. The main focus of attention in this thesis are *nuclear* intonation patterns, i.e. the final sentence accent of an utterance including the phrase-final intonation. Word accents play a further important role, since it is also investigated how intonation patterns are realised in parallel with the word accents. However, less attention is paid to phrasing phenomena.

## 1.3  Outline

The thesis comprises seven chapters in addition to this introduction. In Chapters 2 and 3, the theoretical background of the thesis is described and some earlier work,

primarily on Swedish, is discussed, resulting in a more specific presentation of the research questions. Chapter 4 provides an overview of the data and methods used in this thesis.

Analyses and results are presented in three chapters, comprising three production studies (Chapters 5, 6, and 7.2) and one perception experiment using reaction time measurements (7.3). The first two production studies (Chapters 5 and 6) explore how selected sets of pragmatic distinctions are signalled intonationally in German and Swedish. The first of these studies (Chapter 5) constitutes an initial step in testing the general hypothesis. One finding is that in Swedish, as also in German, a *confirmation* may be signalled by means of a *falling* intonation pattern, phonetically more similar to a *non-focal word accent* (e.g. H+L*) than to a (rising) *focal* accent (H-). It is argued, however, that it is inappropriate to classify such a falling pattern in confirmations as a 'non-focal' variant of the word accent. Instead, a more specific sub-hypothesis is formulated, namely that Swedish has a low (or falling) accent at the utterance level, besides the high (rising) one (H-), implying a *paradigmatic* contrast of accents at the utterance level, similar to that in the West Germanic languages. Chapter 6 provides a first test of this hypothesis and investigates whether *focus* – understood as a *semantic* notion (e.g. Krifka 2007), as defined in 2.3.1 – can be signalled by a falling pitch pattern in confirmations.

Since on formal grounds, the falling pattern found in confirmations seems to be similar to a non-focal word accent I pattern, the last two studies (both presented in Chapter 7) are concerned with the interplay of utterance- and word-level prosody. Thus, the signalling of *confirmation* and its interplay with the signalling of other prosodic functions (focus and word accent), is a central theme of the thesis and connects all four sub-studies. The results are summed up and discussed in a concluding Chapter 8.

# Chapter 2

# Intonation and its functions

This thesis operates with some notions which on the one hand represent basic concepts in prosody research, but which on the other hand have been used with sometimes crucially different interpretations in the literature. As mentioned in Chapter 1, the thesis concentrates on phenomena related to *accentuation*, paying less attention to *phrasing*. First and foremost the terms *intonation*, *(pitch) accent*, and *focus* need to be handled with some care when it comes to a comparison of terminology used in the German and the Swedish traditions. One goal of this section is therefore to try to clarify the most problematic notions in order to develop a common terminological ground for the discussion of German and Swedish prosody, with focus on phenomena related to accentuation.

Investigating and modelling prosody and intonation implies the task of pinpointing form–function relations: What are the units of prosody and what are their communicative functions or meanings? As mentioned in 1.1, the point of departure of this thesis is the observation that different research traditions or perspectives have been underlying the majority of studies on German and Swedish prosody. It is suggested that a major difference in perspective between the German and the Swedish tradition is related to the *functions* of intonation. Therefore, as a second goal, this chapter also attempts to provide a brief overview of functions and meaning of intonation in general. Special attention, however, is paid to areas that are most relevant for the discussion of German and Swedish in chapter 3, and the studies outlined in 3.5.

# 2.1 Some basic notions

## 2.1.1 Intonation and prosody

In two typical definitions of *prosody*, the term refers either to the *melodic*, *rhythmical* and *dynamic*, or to the *suprasegmental* aspects of language and speech (e.g. Bruce 1998). While probably neither of these two definitions is entirely satisfying (cf. Bruce 1998 for a discussion), they are nevertheless largely compatible, and each of them is sufficient for the present thesis. Prosody, in both definitions, implies a variety of rather different phenomena, such as abstract phonological features like lexical stress, but also pausing phenomena, or the paralinguistic usage of pitch or voice quality. The two prosodic phenomena which are investigated in this thesis are *intonation* in German and Swedish, and the Swedish *word accents*.

Before discussing these notions, we will briefly treat the phonetic correlates of prosody. The impressionistic labels used in the definition of prosody above (i.e. *melodic*, *rhythmical*, and *dynamic*) are typically related to the psycho-acoustic dimensions *pitch*, *length*, *loudness*, and often also *timbre* (e.g. Jones 1969).[1] These psycho-acoustic notions have correlates in the acoustic domain, most importantly *fundamental frequency* ($F_0$), *duration*, *intensity*, as well as the distribution of energy in the *sound spectrum*. Finally, these acoustic features have correlates in speech production, such as the vibration of the vocal folds or the timing of laryngeal and supra-laryngeal articulatory gestures.

However, there is no one-to-one mapping between these different levels of description. For instance, the perception of *melodic* patterns has been shown to integrate both $F_0$ and spectral information (House 1990). Another example is that $F_0$ is not only the result of active laryngeal control, but also influenced by myoelastic and aerodynamic interactions with supra-glottal articulation, which give rise to so-called *micro-prosodic* effects.[2]

The term *intonation* is sometimes used broadly, referring to the same phenomena as *prosody* (e.g. Baumann 2006), but commonly, the term is restricted to the *melodic* aspects of utterances (e.g. Jones 1969; Selting 1987; 't Hart et al. 1990). *Speech melody* can be characterised as the "temporal organisation of perceived pitch of utterances" (Selting 1987), although Niebuhr (2007b) also integrates *loudness* in the definition of speech melody. However, pitch and its tem-

---

[1]The prototypical usage of timbre in speech is not related to prosody, but to the distinction between different vowels. However, since vowel quality can vary depending on stress, some changes in timbre can be regarded as a prosodic phenomenon.

[2]For instance, $F_0$ is typically higher in close than in open vowels (e.g. Lehiste and Peterson 1961; Reinholt Petersen 1978; Antoniadis and Strube 1981; Ladd and Silverman 1984), and lower during voiced obstruent as compared to sonorant consonants (e.g. Laver 1994). $F_0$ is also higher after the release of a voiceless compared to a voiced consonant (e.g. Löfqvist et al. 1989).

poral organisation can be regarded as the primary psychoacoustic correlates of speech melody. Hence, in this thesis, as in most comparable empirical studies, intonation (and word accents) are investigated by means of analysing $F_0$, and to some degree duration patterns.

Definitions of intonation differ concerning the 'linguistic scope' of the term. While intonation has often been used to refer to *all* melodic features of utterances, it is also common to restrict the term to those melodic properties that have communicative functions related to the sentence- or utterance-level, hence excluding the word-level usage of pitch such as lexical tone (e.g. Lehiste 1970; Haugen and Joos 1972; Kohler 1995; Gussenhoven 2004; Ladd 2008[3]). This restriction of the term is adopted here, and hence,

- *intonation* refers to the post-lexical component of *speech melody*, i.e. to those aspects of *speech melody* that have a communicative function related to the phrase, sentence, utterance, or the discourse.[4]

That is, Swedish word accents, although melodic, are not regarded as a part of Swedish intonation, since they are lexically or morphologically determined and not related to utterance- or discourse-related functions. However, all other pitch-related accents (cf. 2.1.2), such as the Swedish *focal accent* (cf. 2.1.2), or any accent of German, belong to the intonational system of the corresponding language. Hence, by referring to *Swedish intonation* it is intended to exclude the word accents. When the latter are to be included, terms such as *tonal* or *melodic* aspects of Swedish may be used. The adjectives *tonal* and *melodic* are used interchangeably in this thesis.[5]

In the Swedish context, *intonation* has been used with both scopes, either including or excluding the word accents. For instance, when referring to a Swedish *intonation model* (e.g. Bruce 1982a), this would typically account also for the word accents. This usage of the phrase *intonation model* can also be adopted here, even if the word accents are excluded from the notion of *intonation*, since an *intonation model* should, of course, also account for the word accents and their interaction with intonation.

---

[3]However, Ladd's (2008) definition is actually narrower than the more traditional usage adopted in this thesis, since it excludes any paralinguistic variation (i.e. *gradual* as opposed to *categorical* variation) from the notion of intonation.

[4]The crucial distinction made in this thesis is between the word level and any higher level, but distinctions between e.g. the phrase and the utterance are not relevant here. Terms such as *phrase-level* and *utterance-level phenomenon* are thus used interchangeably in this thesis.

[5]Hence, *tonal* is not used to refer to *lexical tone* only, although the noun *tone* is often, as in Chapter 1, used in this restricted sense.

## 2.1.2   Prominence and accentuation

*Prominence* refers to the *weighting* (*highlighting* vs. *withholding*) of linguistic or phonetic units (Bruce 1998). That is, a *prominent* unit is 'more salient' than at least one other unit of its kind. Such a unit may e.g. be a syllable, a word, or even a whole phrase or sentence.

Prominence has both a concrete phonetic (i.e. acoustic and perceptual) and an abstract phonological dimension. In the phonetic sense, "prominence is the property by which linguistic units are perceived as standing out from their environment" (Terken 1991). For example, a word uttered with a salient movement in $F_0$ will usually be perceived as more prominent than other words in the utterance with only a slight $F_0$ movement or flat $F_0$. Increasing the $F_0$ range of the excursion will render the word even more salient, and hence more prominent. Thus, in this first sense, prominence is *gradual*: A unit is more or less prominent.

In the phonological sense, prominence is *categorical*: A unit is either prominent or not, but there can also be distinct prominence *levels* in a *prosodic hierarchy* (as exemplified below). For example, in languages that have *lexical stress*, like German and Swedish, the stressed syllable of a word is associated with an abstract prominence, while the others are not (a brief introduction to lexical stress in German and Swedish is provided in 3.1). This prominence is abstract, because it is a feature of the word which is present whether the word is uttered or not (i.e. speakers of a language know which syllable is the stressed one in a word). This abstract, phonological prominence can become concrete when the word is uttered, especially when the word is associated with a *pitch accent*. Lexical stress in languages like German, English, and Swedish, is typically regarded as the docking site for a pitch accent.[6]

The term *pitch accent* can receive different interpretations, which is mainly due to the fact that pitch accents can serve different functions, or that they can signal prominence within different *domains*, as discussed in what follows. A first basic distinction can be made between *lexical pitch accents* and *intonational pitch accents*. Ladd (2008) provides a definition of pitch accent, and states explicitly that his definition is intended to account for intonational pitch accents only. However, a slight modification of his definition will make it more general:

- "A pitch accent is a local feature of a pitch contour – usually but not invariably a *pitch change*, and often involving a local maximum or minimum

---

[6]However, even in non-accented contexts, lexically stressed syllables typically have some phonetic prominence e.g. through increased duration (Sluijter and van Heuven 1996; Bruce 1998; Gussenhoven 2004), a more even energy distribution in the spectrum (Sluijter and van Heuven 1996; Sluijter et al. 1997) or even a slight $F_0$ movement (Xu and Xu 2005), as compared to unstressed syllables.

– which signals that the syllable with which it is associated is *prominent*"
(Ladd 2008) in a *certain domain*[7] [ga].

As indicated by the definition and the term itself, pitch (or $F_0$) is the predominant psychoacoustic (or acoustic, respectively) correlate of a pitch accent, although there are further correlates such as increased duration (e.g. Cambier-Langeveld and Turk 1999).[8] The definition presented here is intended to account for both intonational pitch accents found in German and Swedish and the lexical pitch accents found in Swedish, which are also referred to as *word accents*. Hence, first, a pitch accent can highlight a syllable at the utterance (or phrase) level, or within the *domain* of the utterance (phrase). This kind of *sentence accentuation* (cf. below) exists in both German and Swedish. A typical purpose of this sentence- or utterance- level highlighting is to signal that the word which accommodates the accent has some special information status in the discourse (as discussed in more detail in 2.2.2, 2.3).

Second, in addition to a sentence accent, Swedish also has lexical pitch accents, or word accents, which are not determined by sentence-level function, but by the lexicon or morphology, as discussed in 3.3.1. However, according to Bruce (e.g. 1977, 1998), the Swedish word accents also signal prominence, although within another domain in the prosodic hierarchy, namely the domain of a *prosodic word*.

Table 2.1 illustrates the prosodic hierarchy, or the structure of prominence levels, assumed by Bruce (1977, 1998; Bruce and Hermans 1999) for Swedish. Bruce (e.g. 1998) assumes three relevant prominence levels, besides a fourth *unstressed* category. The three levels are related to the three prosodic domains *foot*, *prosodic word*, and *phrase*. At the lowest level, syllables can be (lexically) stressed or unstressed. Syllables are grouped into feet, and each foot contains one stressed syllable. In the example of Table 2.1, there are five syllables which are grouped into three feet. Words (i.e. prosodic words) can comprise several feet (e.g. two in the prosodic word *lamadjur på* [ˈlɑː ma ɟʉːr pɔ] in the example), and hence several stressed syllables. One stressed syllable in each prosodic word is the primary stress and associated with a word accent. That is, word accent is assumed to signal that the foot that contains the primary stress is prominent within the word.[9]

Hence, each word in Swedish contains one or several feet, one of them being associated with a word accent, and each phrase contains one or several (prosodic)

---

[7]In Ladd's (2008) definition "in a certain domain" is originally "in the utterance".

[8]Besides pitch accents, *force accents* have also been reported on (Kohler 2005a; Kohler and Niebuhr 2007). *Force accents* can take the form of 'non-pitch' accents.

[9]There is a choice between accent I and accent II (discussed in more in detail in 3.3.1), but this choice does not affect the prominence of the accented foot.

Table 2.1: The structure of prominence levels in Swedish, exemplified with the phrase *lamadjur på Zoo* 'llamas in the Zoo'. ['] indicates the sentence accent. Parentheses mark domain boundaries. The '.' symbolises unstressed syllables. Modified from Bruce and Hermans (1999).

| | | | Example | | | | |
|---|---|---|---|---|---|---|---|
| | Prosodic category | Domain | [ˈlɑ̀ː | ma | ɟɰːr | pɔ | ˈˈsuː] |
| 1. | stress | foot | (+ | . ) | (+ | . ) | (+) |
| 2. | word accent | prosodic word | (+ | | | ) | (+) |
| 3. | sentence accent | phrase | ( | | | | +) |

words, one of them being associated with a sentence accent.[10]

For Swedish, it is therefore convenient to speak of two distinct *levels of tonal prominence* (Bruce 1998) related to accentuation within two different domains in the prosodic hierarchy (the prosodic word, the phrase). However, the term *accent* has been used ambiguously in the Swedish tradition. On the one hand, it applies to both levels of accentuation (cf. the common terms *word accent* and *sentence accent*), but on the other hand, it has also been used in a restricted sense applying to word accents only. For instance, Bruce (1998) refers to the two levels of prominence as *accent* and *focus*. In this thesis, the term accent is only used in the wider sense, applying to both word and sentence accents.

The term *pitch accent* is quite common in research on English and German intonation, but less so in the Swedish tradition.[11] Instead, it is more common to simply use the notion *accent* and to modify it according to its function as *word accent* on the one hand, and the *sentence accent*, or *focal accent*, on the other hand (cf. also 2.3.1).

Also in the German tradition, the term *sentence accent* (or "Satzakzent") has been common (e.g. Kohler 1995), in order to distinguish (sentence) accents from lexical stress. That is, the term *sentence accent* has been used both in the German and in the Swedish tradition to refer to *intonational pitch accents* (all accents of German and the focal accent of Swedish), and can hence be adopted with this meaning in this thesis. As mentioned in Chapter 1 and discussed in more detail in Chapter 3, for Swedish, only one type of sentence accent has been assumed,

---

[10]Other notions used for the concept of a *sentence accent* (Bruce 1977) are the *phrase accent* (Bruce 1982a) or the *focal accent* (Bruce and Granström 1993).

[11]Similar terms such as *tonal accent* have been used (e.g. Elert 1964), however, in a more special sense, namely referring to accent II, i.e. *one* of the two word accents according to Table 1.1 (cf. the discussion of the Swedish word accents in 3.3.1).

which has been referred to as 'the' *sentence accent* (e.g. Bruce 1977), and later
the *focal accent* (e.g. Bruce and Granström 1993). While *sentence accent* is used
in this thesis as a general term for accents at the phrase (or utterance) level, the
term *focal accent* can be reserved for the specific sentence accent H- (cf. Table
1.1 in Chapter 1; discussed in more detail in 3.3.3) modelled for Swedish.

Finally, a notion relevant in connection with prominence and accentuation
is the *nucleus*. This notion goes back to the British school (e.g. Crystal 1969;
O'Connor and Arnold 1973) and introduces a distinction between *nuclear accents* – defined as the most prominent and last accent in an intonation phrase –
and *pre-nuclear* accents, i.e. any accents preceding the nuclear accent. The nucleus concept is common in intonation models for German and English, but uncommon in the Swedish tradition (cf. 3.4). As discussed in more detail in later
sections (2.2.1, 2.2.3, 2.3.1, 3.2.3), the nuclear accent is typically assumed to be
the functionally most important accent in an utterance, since it can be associated
with different melodic patterns and contribute different meanings to the utterance.

## 2.2   Classifying functions and meanings of prosody and intonation

The functions of prosody and intonation can be described and classified at different levels. Examples of the most general descriptions can be found in Barry
(1981) or Bruce (1998). According to Bruce (1998), prosody has two fundamental functions – (i) signalling *prominence* (highlighting and withholding; cf.
2.1.2) and (ii) *grouping* (signalling coherence and boundaries) – plus a pool of
various functions, such as the signalling of speech acts, which Bruce (1998) summarises as (iii) *discourse-related* functions. Bruce's (1998) classification accounts
for prosody in general, since e.g. prominence relates both to intonation (sentence
accents), and to the word level (word accents and stress).

Barry (1981) recognises five basic communicative functions of prosody and
two fundamentally different types of functions, as illustrated in Table 2.2. The
basic difference between the *intrinsic* and the *relational* functions is related to the
context dependency of the function. According to Barry (1981), the *intrinsic* functions "operate on the basis of the prosodic structure alone", while the *relational*
functions "depend on the integration of prosodic, syntacto-semantic, contextual
and situative information" (Barry 1981).

The *guide* function refers to the fact that prosody helps the listener to extract
the relevant signal from noise and to direct the listener's attention to the most
important elements of an utterance. This is achieved by the "global rhythmic
and tonal pattern", which can, for instance, aid the listener to predict upcoming

Table 2.2: Basic prosodic functions according to Barry (1981). See text.

| Prosodic function | | | | |
|---|---|---|---|---|
| Intrinsic | | | Relational | |
| *guide* | *delimitation* | *focussing* | *interactive* | *attitudinal* |

accents. This function is hence a very general function and related to both of Bruce's (1998) fundamental functions *prominence* and *grouping*.

The *delimitation* function is also related to the "global rhythmic and tonal pattern", but it corresponds more directly to Bruce's (1998) *grouping* function. The *focussing* function is related to accentuation, and approximately to Bruce's (1998) *prominence*. A crucial difference, in both cases, is that Bruce's (1998) functions are intended to be of a more general kind, since they apply also to the word level, while Barry (1981) concentrates on the utterance level. Barry's (1981) *focussing* function, for instance, is rather specific, since it refers to the localisation of the main accent within a phrase (a "sense unit") to determine the "focus of information" (Barry 1981). Furthermore, it has a special, intermediate status between the two basic types of functions: it is not entirely intrinsic, but not entirely relational, either. This characterisation is well in line with the notion of *focus* as discussed in 2.3.1 below. In short, there is a close link between accent placement and focus location (hence the intrinsic character of the *focussing* function), but this relation depends also on the context (as evident e.g. in the distinction between *broad* and *narrow focus*, discussed in 2.3.1). Moreover, different *pragmatic* (i.e. context-dependant) *usages of focus* can be distinguished (2.3.1).

In general, while Barry's (1981) intrinsic functions (the *guide*, the *delimitation*, and some aspects of the *focussing* function) seem to correspond to Bruce's (1998) two basic functions, *prominence* and *grouping*, the relational functions (the *interactive*, the *attitudinal*, and some aspects of the *focussing* function) correspond to Bruce's (1998) *discourse* functions. Barry's (1981) *interactive* function is, for example, related to turn taking and other discourse phenomena and signalled by the "direction of pitch movement" and the "general pitch level". Finally, speaker attitude (the *attitudinal* function) is signalled by parameters such as "type of pitch movement" and "pitch level and pitch range" (Barry 1981).

In order to reach a comprehensive understanding of prosodic form–function relations both from language-specific and cross-linguistic perspectives, general classifications of prosodic functions of the kind exemplified here so far, need to be complemented with more specific descriptions or models. But before continuing the discussion of prosodic functions in more detail (Sections 2.2.2, 2.2.3,

2.4), Section 2.2.1 considers some structural properties of intonation – a *syntagmatic* and a *paradigmatic* dimension – and discusses how these properties could be related to types of prosodic functions.

## 2.2.1   Syntagmatic and paradigmatic aspects of prosody

Prosody can be conceived of as being structured along two dimensions, a *syntagmatic* and a *paradigmatic* one. Gussenhoven (2004), for instance, mentions these two dimensions in connection with lexical tone. The paradigmatic dimension refers to the number of tonal contrasts possible on a given syllable, while the syntagmatic dimension refers to the number of positions in a word where these contrasts are used.

This two-dimensional view of prosody can also be applied to intonation (e.g. Halliday 1967a; Peters et al. 2005). At a syntagmatic level, utterances can differ in the placement of accents and boundaries. It would thus seem that Barry's (1981) *intrinsic* functions of prosody, or Bruce's (1998) two basic prosodic functions *prominence* and *grouping*, are closely related to the *syntagmatic* dimension of intonation.

At the *paradigmatic* level, utterances can differ in the *type* of accent or the *type* of boundary in a given position. For example, for many languages, including German, English, and Swedish, a paradigmatic contrast of *final boundary tones* – high vs. low – has been suggested (cf. Table 1.1 in Chapter 1). The difference is not related to phrasing as such, since any of the boundary tones necessarily implies a phrase boundary at that position, but rather with interactive (e.g. question vs. statement) or attitudinal (e.g. friendliness) functions, as discussed in more detail in 2.4. Similar observations can be made for paradigmatic contrasts of *accents*, or the entire *nuclear pattern*[12] in languages like English and German, as exemplified below (2.2.3, 3.2.3). It would seem that the *paradigmatic* dimension of intonation is closely related to Barry's (1981) *relational* functions, or Bruce's (1998) various *discourse* functions of prosody.

It is not claimed here that there is a one-to-one mapping between the two dimensions of intonational structure on the one hand and the basic distinction of intrinsic vs. relational functions on the other hand. However, for the phenomena touched upon in this thesis, it appears to be adequate to operate with the notions of a syntagmatic vs. a paradigmatic dimension, and to relate them to the classification of functions as intrinsic vs. relational.

---

[12]*Nuclear pattern* is used to refer to the *nuclear accent* (cf. 2.1.2) plus the melodic pattern following the nuclear accent until the end of the intonation phrase.

### 2.2.2 Intrinsic functions and the syntagmatic dimension of intonation

As suggested above, Barry's (1981) *intrinsic* functions of prosody are related to Bruce's (1998) two fundamental prosodic functions, the signalling of *prominence* and *grouping*. The discussion here is restricted to intonation, which in the case of prominence means that we are concentrating on sentence accents, leaving aside lexical stress and word accents.

Sentence accents can be used to lend *prominence* to words at the utterance level. A common purpose of this is to highlight certain pieces of information as important in the discourse; this is discussed in more detail in connection with *information structure* in 2.3.

As for the *grouping* function of prosody, speech is produced in chunks, or *phrases*, of varying length, which may, but need not coincide with syntactic units. The typical functions that are associated with this chunking are to aid both the listener and the speaker in the perception and the production of speech (e.g. Hansson 2003; Horne et al. 2006). Phrasing can also disambiguate syntactic structures and "indicate which words within a sentence belong together semantically or pragmatically" (Hansson 2003). The grouping function implies thus both the signalling of phrase *boundaries* and the signalling of *coherence* of constituents, e.g. words within phases but also phrases within a larger utterance. This thesis is primarily concerned with accentuation, rather than with phrasing. However, phrase-final intonation, as a component of *nuclear patterns*, as well as the signalling of coherence are also touched upon.

### 2.2.3 Relational functions and the paradigmatic dimension of intonation

It has been suggested in 2.2.1 that paradigmatic intonational contrasts are related to *relational* (i.e. attitudinal and interactive) functions of intonation in the sense of Barry (1981), or to *discourse-related* functions in the sense of Bruce (1998). For instance, the choice of a certain final boundary tone can, in some contexts, code a distinction between a statement (L%,[13] i.e. falling intonation) and a question (H%, i.e. rising intonation). To give a further example, by the choice of either of the pitch accents H* (a pitch peak timed medially in the stressed syllable) or H+L* (an early-timed pitch peak) in German (cf. 3.2.3 below), the speaker can express that s/he either presents the information which is associated with the accent as 'open for discussion' (H*) or as an 'unchangeable fact' (H+L*); a third type of accent (L*+H) can imply surprise (e.g. Kohler 1991a). The choice of accent may also

---

[13]These AM-based tonal labels are introduced in more detail in Chapter 3.

be relevant for signalling different *degrees of givenness* of information (Baumann 2006) or to signal *focus* in different contexts, which is discussed in more detail in 2.3.1. That is, the relational functions of intonation, related to paradigmatic intonational contrasts, appear to represent a rather heterogeneous set of functions or meanings.

There seems to be no consensus in the literature about the distinction, or the synonymy, of the notions of *function* and *meaning* in connection with intonation. In fact, this issue is seldom discussed explicitly, but, implicitly, some authors distinguish between the notions (e.g. Kügler 2007; Niebuhr 2007b)[14]. However, there seems to be a tendency that the concept of *meaning* is mainly used in studies concerned with relational functions, rather than with intrinsic functions. That is, authors usually do not claim that "the *meaning* of an L% boundary tone is *boundary*", but rather that "the presence of the boundary tone has the *function* of signalling a boundary". However, it is more natural to say that the "the meaning (or the function) of the L% is different from the meaning (or the function) of the H%".

Although a variety of opinions exist as to what the distinctive and meaning-ful units of intonation are, there is a wide consensus among authors concerned with phonological aspects of intonation that "the elements of intonation have morpheme-like meaning" (Ladd 2008), an assumption that Ladd (2008) regards as the central idea in the "Linguist's Theory of Intonational Meaning".

A common assumption concerning such morpheme-like units of intonation is that their meanings must be of a very general kind. That is, the meaning of an intonational morpheme is usually not as concrete as the meaning of words. Intonational meanings, instead, receive a concrete pragmatic interpretation when put into context, and the number of possible concrete interpretations is large.

First and foremost for English, some serious proposals have been made for general, underlying meanings of certain units of intonation. For instance, Gussen-hoven (1984) regards three basic nuclear contours of English – the fall, the rise, and the fall-rise – as morphemes with general meanings referred to as *addition*, *relevance testing*, and *selection*.

Pierrehumbert and Hirschberg (1990) propose a more detailed system. Both accounts are based on AM phonology and hence formally compositional, i.e. con-tours are composed of tones. The most important difference between Pierrehum-bert and Hirschberg's (1990) and Gussenhoven's (1984) approaches is that the former not only regard intonational *form*, but also its *meaning* to be composi-tional. That is, each tone has a basic meaning component, and these components

---

[14]Kügler (2007) uses both notions without explicitly explaining the distinction, while Niebuhr (2007b) only refers to *meanings*. Both, however, distinguish between the concepts (personal com-munication).

are combined in pitch accents or nuclear patterns where several tones are included. Gussenhoven (1984), instead, assumes that tones have a distinctive function but no meaning, and that basic meanings can only be defined for larger tone complexes (the *nuclear patterns*), his morphemes.[15] Both approaches, however, fit Ladd's definition of a "Linguist's Theory of Intonational Meaning".

However, attempts to pinpoint underlying, context-independent meanings, as the ones by Gussenhoven (1984) and Pierrehumbert and Hirschberg (1990), are rather infrequent, an exception for German being, for instance, Niebuhr (2007b). Instead, many descriptions or models of intonation only provide context-dependent interpretations in order to exemplify the meanings of their proposed intonation patterns. This is exemplified in more detail for the case of German in 3.2.3. As argued in chapter 3, relational meanings have played a minor role in Swedish intonation research so far.

## 2.3   Information structure

One function of intonation that has received much attention is the signalling of *information structure* (IS), a term going back to Halliday (1967b). IS can be understood as "the partitioning of sentences into categories such as focus, background, topic, comment etc." (Büring 2007). Even if the signalling of IS is not a matter of intonation alone, the formal marking of IS makes extensive use of intonation, since it involves both syntagmatic (related to the placement of accents) and paradigmatic (type of accents) intonational means.

IS research involves a wide range of disciplines, such as phonology, syntax, semantics, and pragmatics. This thesis is mainly concerned with phonological and phonetic aspects of intonation, but some of the intonational form–function relations investigated in this thesis are related to IS phenomena. Hence, it is neither possible to give an exhaustive account of IS here, nor is it necessary, since only a small selection of IS-related phenomena are relevant for the issues of intonational phonology treated in this thesis. This selection comprises *focus*, including different *pragmatic usages* (Krifka 2007) of focus as discussed below, and *contrastive topic*.

The process of transmitting information during communication implies that a sender adds some relevant information to the knowledge of a receiver. Two basic aspects of communication are involved in this process (cf. Hetland and Molnár 2001): First, the sender has to estimate the knowledge of the receiver in order

---

[15]An argument for Gussenhoven's (1984) contour-based approach of intonational meaning is provided e.g. by Kügler's (2007) analysis of two dialects of German. According to Kügler (2007), both dialects have three nuclear patterns corresponding to the three basic meanings proposed by Gussenhoven (1984), but for one of the dialects, each nuclear pattern contains a low pitch accent.

to choose an appropriate point of departure for her/his utterance. Second, the sender has to signal what s/he wants to be understood as the relevant information to be added to the receiver's knowledge. That is, the sender has to ensure that the utterance is *coherent* with the discourse as well as *informative* (Hetland and Molnár 2001). Hence, basically all accounts of information structure assume a division of a sentence into (at least) two basic units corresponding to the *coherence* aspect and the *information* aspect. A common notion related to the coherence aspect is *topic*,[16] while the information aspect is related to *focus*.[17] In a tentative approach, the *topic* can hence be understood as the 'starting point' of an utterance, connecting it with the preceding discourse, while the *focus* is the 'informative part'. These notions are exemplified in (1).[18]

(1)     A:   What did your brother give you for your birthday, actually?
        B:   [That stupid guy]$_{Topic}$ gave me [a COOK book]$_{Focus}$.

The example also provides a first impression of the importance of intonation for IS signalling: The focus is typically associated with a sentence accent in languages such as English, German, or Swedish, while the topic can, but need not be accented. A special type of topic which is generally accented is a *contrastive* topic as discussed in 2.3.2.

A further parameter of IS is *givenness*, or the distinction of *new* vs. *given* information. *Focus* is often related to the *new* information in a sentence, while *topic* refers to *given* information. Example (1) is a typical illustration of this correlation, at least if we assume that A understands that B is using the phrase *that stupid guy* in order to refer to his brother.

However, it is important to recognise that there are different *types of focus* and different *types of topic*, as discussed in more detail in the following sections, which take a closer look at the notions of *focus* (2.3.1) and *contrastive topic* (2.3.2). To anticipate the most relevant issues for this thesis, the *topic* and *focus* structure of a sentence is closely related to, but principally independent of the information status (*new–given*) of discourse referents. Moreover, even topics can be accented, and typically a different type of accent is used in a contrastive topic than in focus, at least in languages like German and English. Different types of accents may also play a role for signalling different *pragmatic usages* of focus.

---

[16]Related notions are *background*, *presupposition*, and *theme*.

[17]Related notions are *comment* and *rheme*.

[18]In this and all following examples in this thesis, the relevant sentence accents are indicated by capitals. Optional, additional accents whose presence would not change the IS of the sentence are not indicated.

## 2.3.1 Focus

*Focus* has received a variety of definitions. The traditional definition of focus as the *new information* goes back to early understandings of focus such as the one in Halliday (1967b), according to which 'focus reflects the speaker's decision as to where the main burden of the message lies'. Halliday (1967b) himself paraphrased this definition in terms of *new information*.[19]

As exemplified below and argued by e.g. Krifka (2007), the understanding of focus as new information is "intuitively appealing and may apply to a majority of cases [...] But it clearly gives us wrong predictions. There are many cases in which a constituent that refers to something mentioned previously is in focus" (Krifka 2007). Krifka (2007) provides an alternative, more general definition based on Rooth (1992). Krifka's (2007) definition of focus is adopted for this thesis:

- "*Focus* indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions".

The definition of focus as new information (e.g. Halliday 1967b) may be regarded as a *pragmatic* understanding of focus, because it relates focus to contextual, cognitive, or attitudinal factors. The definition of focus as an indicator of alternatives, on the other hand, has a *semantic* character. This semantic dimension of focus is primarily discussed in works on so-called *focus-sensitive particles* such as *only*, *also*, and *even* in English (e.g. Rooth 1992; Krifka 2007).

These two aspects of focus (the *semantic* and the *pragmatic* one) have already been implied in Jackendoff's (1972) account of focus[20], but the distinction between the two aspects has been comprehensively discussed for the first time by

---

[19]"Information focus is one kind of emphasis, that whereby the speaker marks out a part (which may be the whole) of a message block as that which he wishes to be interpreted as informative. What is focal is 'new' information; not in the sense that it cannot have been previously mentioned, although it is often the case that it has not been, but in the sense that the speaker presents it as not being recoverable from the preceding discourse. The focal information may be a feature of mood, not of cognitive content, as when the speaker confirms an asserted proposition; but the confirmation is itself still 'new' in the sense intended." (Halliday 1967b)

[20]In Jackendoff's (1972) working definition, *focus* is "the information in the sentence that is assumed by the speaker not to be shared by him and the hearer". That is, on the one hand, Jackendoff's (1972) understanding of focus is similar to Halliday's (1967b), since it can also be paraphrased as 'focus is the new information'. However, a key idea in Jackendoff's (1972) division of a sentence into focus and presupposition is that, somewhat simplified, the focus contains the semantic material that is not part of the presupposition; the presupposition, in turn, is derived from a semantic variable which represents a "coherent class of possible contrasts with the focus, pieces of semantic information that could equally well have taken the place of the focus in the sentence, within bounds established by the language, the discourse, and the external situation." That is, on the one hand, the idea of set of alternatives is present, while on the other hand, this set is shaped by the context.

É. Kiss (1998), who distinguishes between two types of focus, an *identification* focus (related to the semantic aspect), and an *information* focus (related to the pragmatic aspect).

However, as pointed out by Krifka (2007), focus, as a *semantic* notion, also has an additional *pragmatic* dimension, since the choice between alternatives is always related to a certain context. In Krifka's (2007) terms, there are different *pragmatic usages* of focus, the classical example being to highlight the part of an answer that corresponds to the *wh*-part of a constituent question. In this pragmatic usage of focus, the focussed information is typically *new*. However, depending on the relation between the *context* and the *set of alternatives*, the pragmatic usage of focus can also receive e.g. a *corrective* or a *confirmative* character, as exemplified below. Hence, Krifka's (2007) proposal implies both relevant dimensions of the concept of focus (the semantic and the pragmatic one).

To conclude, the term *focus* is here understood as a *semantic* notion, which can have different *pragmatic* usages, following Krifka (2007). The overview presented in the following sections discusses this understanding of *focus* in more detail, but concentrates on some properties of focus which are especially relevant in the light of this thesis. First, focus has a syntagmatic dimension, i.e. focus can have different domains as signalled by the specific placement of accents. Second, as already mentioned, *focus* is conceptually distinct from the *given–new* distinction. Third, there are different *pragmatic usages* of focus, some of which will be presented below based on the account of Krifka (2007). It will be suggested that these different *usages* of focus might be related to the paradigmatic choice of accent type.

**Focus and accent placement**

Focus involves highlighting of the focussed material. This highlighting can be achieved by different formal means, which can involve syntax, morphology, or phonology, including prosody (e.g. Gussenhoven 2007). Languages differ in their choice and combinations of the different formal options. For instance, *prosodic* highlighting is well-known to be common in Germanic languages (e.g. German, Dutch, English, Swedish).

Prosodic highlighting is usually achieved by means of accents, but there is no one-to-one mapping between an accent and focus, as originally suggested by Bolinger (e.g. 1985). This is exemplified in (2) to (4).

(2)      A:    What did you learn at school today?
         B:    [Bears eat BERRIES]$_{Focus}$.

(3)      A:    What do bears eat?
         B:    Bears eat [BERRIES]$_{Focus}$.

(4)   A:   What's the relation between bears and berries?
      B:   Bears [EAT]$_{Focus}$ berries.

The examples illustrate that an accent can mark focus domains of different size. The term *broad focus* is common to refer to cases like (2), while *narrow focus* refers to smaller focus domains, which can consist of a single word such as in (3) or (4). In a broad focus, on the one hand, a single accent can mark a focus on a whole phrase. On the other hand, a phrase in broad focus can also contain additional accents (Uhmann 1991). For instance, in the broad focus case in (2), it would be quite natural in languages like English and German to have an additional, pre-nuclear accent on *bears*, without rendering *bears* and *berries* two individual foci.

Accents that signal focus have sometimes been referred to by terms such as 'phonetic focus' or 'focus accent'. An example is the Swedish tradition, where Bruce's (1977) *sentence accent* has been relabelled *focal accent* in later works (e.g. Bruce and Granström 1993). Such a terminology seems plausible, provided the major (or in the best case, the only) function of that accent is the signalling of focus. However, the terminology becomes vague when a term like *focal accent* is abbreviated as *focus*. In the Swedish tradition, *focus* has also been used to refer to the highest level of *prominence* (e.g. Bruce and Hermans 1999, cf. also 2.1.2). For the sake of terminological clarity, in this thesis, the term *focus* is not used for a certain kind of *accent* or *prominence level*, but for a semantic notion as defined above, while the term *focal accent* is used for the specific H- accent of Swedish, as defined in 2.1.2.

### Focus: always new information?

In many studies on phonetic or phonological aspects of intonation, the definition of *focus* as *new* information may be acceptable, since such studies often discuss focus in the context of a question–answer paradigm, like in (2) to (4) above, in order to elicit broad or narrow focus, and to vary (narrow) focus position. In such a question-answer context, *focus* typically renders *new information*.

A first problem with this definition of *focus* is that it is not always clear what is meant by *new*. Prince (1981) discusses three crucially different definitions of the *new–given* distinction that have been applied in the literature. In short, *given* information can be understood as the information the speaker assumes is (a) predictable, or recoverable, by the hearer (e.g. Halliday 1967b), (b) salient in the consciousness of the hearer at the time of the utterance (e.g. Chafe 1976), or (c) knowledge shared by the speaker and the hearer (but which the hearer is not necessarily thinking about).

However, it is widely acknowledged that, independently of the definition of

*given–new*, focus can (and often does), but need not correlate with new information. In fact, the *focus–background* and the *given–new* distinctions are often regarded as two different levels of IS (e.g. Molnár 1991, 1998; Féry 1993; Vallduví and Engdahl 1996; Smith 2003; Baumann 2006; Krifka 2007).

A typical example of a focussed discourse referent that is given because it has just been mentioned in the discourse and hence salient in the speaker's and the addressee's consciousness, is (5), taken from Lambrecht (1994).

(5)    A:    Where did you go last night, to the movies or to the restaurant?
       B:    We went to [the RESTAURANT]$_{Focus}$.

Such examples are easily accounted for when focus is understood as an indicator of the presence of alternatives, as defined above (Krifka 2007). Another example provided by Krifka (2007) is the case of a *confirmation* in (6). Here, again, focus is an indicator of alternatives, provided a situation where several other persons could also be considered to have taken the cookie.

(6)    A:    Mary stole the cookie.
       B:    Yes, [MARY]$_{Focus}$ stole the cookie.

On the one hand, it could be argued that even an utterance like B's reply in (6) contains some *new* information, in the sense that the confirmation as such is what speaker A requested, and hence the positive feedback by speaker B is what is *new* (cf. Halliday 1967b or Footnote 19 above). However, if *focus* then is to be defined as *new*, it would seem more appropriate to classify the initial word *yes* as the focus, or alternatively, the whole utterance assuming a *broad* focus. But this analysis would fail to explain the fact that a *narrow* focus is still possible in a confirmation, as demonstrated in examples (7) and (8), inspired by Ladd (2008). Again, in these examples, focus is an indicator of alternatives.

(7)    A:    Did you pay five or six euros? Five, right?
       B:    Yes, [FIVE]$_{Focus}$ euros.

(8)    A:    Did you pay five euros or five pounds? Five euros, right?
       B:    Yes, five [EUROS]$_{Focus}$.

**Pragmatic usages of (semantic) focus and types of accents**

As mentioned above, *focus*, understood as an indicator of alternatives, can have different *pragmatic usages*, depending to the relation between the *set of alternatives* and the *context* (Krifka 2007). In this section, some pragmatic usages of focus are discussed, which are exemplified in (9) to (11), adopted slightly modified from Krifka (2007).

All examples represent a *narrow focus* (cf. above). The relevant parameters for distinguishing the pragmatic usages illustrated in (9) to (11) are, first, the nature of the set of alternatives indicated by the focus *(open vs. closed)*, and second, the referential status of the focus constituent *(new vs. given)*. The set of alternatives can be *open*, implying that the context does not provide any limitation of the set of alternatives, as exemplified in (9). Alternatively, the set can be *closed*, if one or several focus alternatives are proposed in the context, thereby limiting the set of alternatives, as exemplified in (10) and (11). Examples of a choice from a closed set have also been presented in (5) to (8) above.

(9)    'PLAIN' FOCUS: *open set, referential status: new*

    A:   Who stole the cookie?
    B:   [MARY]$_{Focus}$ stole the cookie.

(10)    CORRECTIVE FOCUS: *closed set, referential status: new*

    A:   Who stole the cookie? Peter, right?
    B:   No, [MARY]$_{Focus}$ stole the cookie.

(11)    CONFIRMATIVE FOCUS: *closed set, referential status: given*

    A:   Who stole the cookie? Mary, right?
    B:   Yes, [MARY]$_{Focus}$ stole the cookie.

If the set is closed, the focus alternative mentioned by speaker B can be a different one than the one proposed in the context (i.e. speaker A's question), as in (10), implying that focus denotes a *new* discourse referent, and that the usage of focus has a *corrective* character. Alternatively, if speaker A mentions the same alternative as B, as in (11) focus denotes a *given* discourse referent, and the usage is *confirmative*. In this case, "the situation must be such that other alternatives are under consideration as well" (Krifka 2007), i.e. in the example, that there are other people who could be considered to have stolen the cookie, without necessarily being explicitly mentioned.

The question relevant for intonation is whether different *usages* of focus differ prosodically, i.e. whether focus is signalled by means of different types of accents depending on the type of context. In this connection, the notion of *contrast* and the prosodic marking of *contrastive* cases of focus have been widely discussed (e.g. Molnár 2006), which most readily correspond to our *corrective usage* of focus.

The corpus study by Hedberg and Sosa (2007) has shown, that "while there are systematic correlations between intonation and information structure categories, these correlations are not as straightforward as is suggested in the literature" (Hedberg and Sosa 2007). Although Hedberg and Sosa (2007) mainly refer to a hypothesised distinction between accent types used for *focus* vs. *topic* signalling, their results suggest that no perfect correlations of accent types and pragmatic focus us-

ages should be expected. The findings by Hedberg and Sosa (2007) thus support early proposals according to which 'there is no contrastive accent as such' (Pike 1945) or 'as far as we can tell from the behaviour of pitch, nothing is uniquely contrastive' (Bolinger 1961).

However, even if the correlations between intonation and information structure categories found by Hedberg and Sosa (2007) are not perfect, their results are still in line with the common view that contrastive cases of focus can (optionally) be signalled by prosodic means such as a specific choice of accent type (e.g. Chafe 1976; Ladd 1980; Uhmann 1991; Mehlhorn 2002; Kohler 2006b). For instance, Kohler (2006b) proposes a specific correlation of various pitch accent types (differing in peak timing, discussed in more detail in 3.2.3) and different types of focus, involving different kinds of contrast, for the Germanic languages. Although Kohler's (2006b) notion of focus is different from the definition adopted here, it could be expected based on Kohler (2006b) that another type of accent (a *late peak* accent) would occur in the *correction* as opposed to a (non-contrastive) focus related to an open set, as in (9). Also, based on the hypotheses available for German accent types (cf. 3.2.3), yet another type of accent (the *early peak* accent) could be expected to be used in a *confirmation* (e.g. Niebuhr 2007b).

The studies presented in this thesis (Chapters 5 to 7) concentrate primarily on the distinction exemplified in examples (9) and (11), i.e. the *'plain' narrow focus* and a *confirmative* usage of *narrow focus*, although also a case of a *corrective* usage (10) is included in the initial study (Chapter 5).

## 2.3.2   Contrastive topic

The notion *topic* has been defined in different ways (cf. Molnár 2006 for an overview) and used to refer to different phenomena.[21] *Topic* can here be understood as in Gundel (1985):

- "An entity, E, is the pragmatic topic of a sentence, S, iff S is intended to increase the addressee's knowledge about, request information about or otherwise get the addressee to act with respect to E."

As in the case of focus, there can be different types of topic, and this thesis (Chapter 6) touches upon a type referred to as *contrastive topic* (e.g. Büring 1999). As mentioned above, non-contrastive topics are typically not marked by intonation, cf. example (1). It has, however, also been mentioned in connection with broad focus above, that an optional, pre-nuclear accent in a topic constituent is possible. In a *contrastive topic* (CTOP) an accent is obligatory, as illustrated in example (12).

---

[21]For instance, a distinction can be made between *frame topics* and *aboutness topics* (Jacobs 2001), only the latter being relevant for this thesis.

(12)    A:   What do wild animals eat?

          B:   [BEARS]$_{CTOP}$ eat [BERRIES]$_{Focus}$.

Three features distinguish a CTOP from a focus. First, as typically reported for a variety of languages, the accent associated with a CTOP is different from the accent associated with the focus. For instance, a *fall-rise* is typically reported to be the accent used in CTOP, as opposed to a *fall* used for focus, in English (e.g. Jackendoff 1972). However, as noted above, Hedberg and Sosa (2007) showed in a corpus study that this correlation of accent type and IS category is less straightforward in natural, spontaneous speech. Second, while a sentence can lack a CTOP, every sentence is typically assumed to have a focus (e.g. Gussenhoven 1984; É. Kiss 1998; Smith 2003). That is, a CTOP, early in the sentence, always co-occurs with a focus, later in the sentence. Third the *type of contrast* associated with a CTOP is different from the contrast in contrastive foci (cf. Molnár 2006). While in a contrastive *focus* (e.g. the *corrective* case above), the contrast is generally *exhaustive*, the contrast in a CTOP is *non-exhaustive*. That is, in a contrastive topic, like in a contrastive focus, a set of alternatives is induced and excluded (Molnár 2006). However, a CTOP implies that not *all* members of this set are excluded, but rather that there is *at least one* member for which the predication does not hold (Molnár 2006). Hence, *wild animals* in example (12) is the topic elicited by speaker A's question, but speaker B chooses to make an assertion only about a sub-set of the wild animals, thereby implying a contrast between *bears* and the remaining members of the set. However, since the contrast is not exhaustive, there may still be some other wild animals that also eat berries. This gives rise to an open question or, as Büring (1999) puts it, a 'residual topic'.

For German, a further special feature of the CTOP, as discussed in 3.2.3 below, is that the CTOP also has an indirect influence on the choice of accent type used for the *focus*. In short, the typical intonation of a sentence containing a CTOP and a focus is a so-called *hat pattern* ('t Hart et al. 1990), with a rising accent for the CTOP and a fall for the focus (e.g. Féry 1993; Jacobs 2001), and a high pitch plateau connecting the accents. When embedded in a hat pattern, another type of falling accent (an *early* fall, cf. 3.2.3) is typically used to signal focus as when there is no preceding CTOP in the sentence (in this case, a *medial* fall is common, 3.2.3), although even in a hat pattern, both types of falls are possible. This variation is not always recognised. Kohler (1991a) provides a perception-based explanation for the preference of an early fall in a hat pattern, which is presented in 3.2.3.

## 2.4 Explaining functions of intonation by biological codes

It is today a widely adopted view that many intonational form–function relations observable in the languages of the world are based on a set of universal codes which have developed from biophysical conditions. Gussenhoven (2002) proposes three such *biological codes*: the *effort code*, the *production code*, and the *size code*, which is Gussenhoven's (2002) term for Ohala's (1983, 1984) *frequency code*. According to Gussenhoven (2002), a biological code can be interpreted at an informational level (relating to the message) and at an affective level (relating to the speaker). These interpretations (first and foremost the informational ones) of the codes can be grammaticalised, as exemplified below.

The *size code*, or *frequency code*, is based on the correlation of body size (or larynx size) and $F_0$, which tends to be higher for a smaller larynx containing lighter and shorter vocal folds than for a larger larynx. The code exploits this correlation for the expression of power relations. Typical *affective* interpretations of the size code are to associate high (or rising) pitch with friendliness, submissiveness, or vulnerability, while low (or falling) pitch is associated with dominance, confidence, or aggression. Typical *informational* interpretations relate high (or rising) pitch to uncertainty or request for information, while low (or falling) pitch corresponds to certainty or assertion (Gussenhoven 2002).

A typical grammaticalisation of this informational interpretation is that rising intonation is used in questions and falling intonation in statements. Gussenhoven (2002) refers to patterns like rises in questions, which are in line with the biological code, as 'natural' and 'unmarked' ones. However, grammaticalisation implies that intonation patterns become part of the phonological system of a language, and hence, they can also be affected by language change. As a consequence, grammaticalised form–function relations can still mimic the original, natural relations, but there can also be, and usually are, arbitrary form–function relations in the intonational systems of individual languages. That is, a language can even develop an 'unnatural' intonation pattern when there is some special need or pressure for such a pattern to develop, e.g. in order to enhance a contrast between two similar patterns in the system. Well-attested cases of 'unnatural' form–function relations are questions with falling intonation.

The case of German exemplifies how the informational, grammaticalised interpretation of the *size code* can interact with the ungrammaticalised, affective interpretation. In German, the grammaticalised status of intonation in questions is evident from the well-established textbook statement that there is an interplay between intonation category and morphosyntax: The presence of a question word correlates with falling intonation, while other types of questions are marked by

a rising intonation (cf. 3.2.3). However, this situation is not very robust. Kohler (2005b) found in a corpus of spontaneous speech, that the opposite intonation pattern, e.g. rising intonation in *wh*-questions, or falling intonation is *yes/no*-questions, occurred in 34%, or 21% of the cases, respectively. Kohler (2005b) explains these results in terms of the speaker's orientation towards the information vs. the addressee, depending on the particular discourse setting. That is, on the one hand, a speaker can choose to express friendliness in a *wh*-question by using a rising intonation. On the other hand, in a *yes/no*-question, where the rising pattern is the default, a rise does not imply any special friendliness. However, choosing a falling pattern instead of the rise in a *yes/no*-question implies a higher degree of confidence of the speaker, indicating an expected polarity in the reply of the hearer. Similar results and conclusions are reported for Swedish in House (2005).

The *effort code* is based on the correlation of articulatory effort and $F_0$ excursion size, where more effort is assumed to lead to larger $F_0$ excursions. The code exploits this correlation for the signalling of relative importance. The effort code, too, can be interpreted on an affective and on an informational level (Gussenhoven 2002). Affective interpretations of the code are related to enthusiasm (interest vs. lack of interest), helpfulness, or surprise, while the informational interpretation is related to emphasis (but cf. Kohler 2006b for a discussion of the concept *emphasis*). The typical grammaticalisation of the effort code is accentuation, which is used, among others, for the signalling of *focus* (cf. 2.3.1). Gussenhoven (2002) mentions that in the case of the effort code, there are no known cases where grammaticalisation has resulted in 'unnatural' patterns of the kind reported for the size code. At best, it is possible to find languages or contexts where important information is not marked with an $F_0$ excursion, but there are no known cases of languages that generally mark the 'unimportant' information with large $F_0$ excursions, while leaving the important information unaccented.

Finally, the *production code* is based on the correlation of $F_0$ and the amount of energy available during the production of an utterance. More energy is available at the onset of an utterance, i.e. directly after inhalation, and $F_0$ tends to fall with available energy. The production code exploits this correlation for the signalling of breaks and continuation. The informational interpretation of this code is to relate high pitch at utterance onset with a new topic in discourse, low pitch at the onset with continuation of a topic, and low vs. high offsets with finality vs. continuation. Gussenhoven (2002) suggests that there is no affective interpretation of this code. The production code is widely grammaticalised, e.g. in form of (language-specific) paradigms of phrase-final or initial boundary tones.

To conclude, in Gussenhoven's (2002) theory of intonational meaning, languages are assumed to have both universal and language-specific intonational meanings. These are related to two different components of language: the pho-

netic implementation (universal meanings) and the intonational grammar (language-specific meanings).  The universal (or "natural" or "phonetic") meanings derive from *biological codes*, and the language-specific meanings develop from the universal meanings through grammaticalisation.

Biological codes are based on the effects of physiological properties of the production process on the signal, but they are only metaphors of these physiological conditions.  That is, communication by means of the codes does not require that these physiological conditions are actually created.[22]  It is sufficient to create the effects, which thus are not automatic, but have been brought under control.

## 2.5   Investigating intonation: Form-first or function-first?

It might seem trivial to state that, in order to fully understand intonation, and hence to be able to model it adequately, the final goal of intonation research must be to understand both the formal and the functional aspects of intonation, as well as the link between them. However, approaches to intonation differ crucially in their treatment of form, function, and the connection between the two levels.  There seem to be at least three research perspectives, although it might not always be the case that a particular study adopts only one of these perspectives.

First, it would seem that the typical paradigm underlying a large proportion of intonation research can be characterised as follows. Intonational *forms* constitute a point of departure, for instance based on a corpus of spontaneous speech, but also based on introspection.  Hypotheses on functions are then derived, again either based on introspection or based on empirical observations (e.g. in which contexts in a corpus does a certain pattern occur?), or based on a combination of both. One common way of testing such hypotheses on functions is by means of a return to the level of form, namely in perception experiments, where formal parameters are manipulated.

That is, typically, intonational forms and functions are explored simultaneously, or in a cyclic process. But since intonational form is the point of departure, this account could also be referred to as a *form-first* approach, although a less strict one (to the clarified below). According to Hirst (2005), a common consequence of this research paradigm is that "[i]n many systems of prosodic annotation, perhaps most, the two levels of representation [form and function; ga] are intimately intertwined". This is reflected by the fact that intonational forms are often referred to by means of function-inspired labels such as *continuation rise* vs. *interrogative*

---

[22]At least in case of the *size code*, this would be physically impossible: we cannot reduce or enlarge the size of our larynx to manipulate $F_0$.

*rise*. Another classical example, as already mentioned in 2.3.1, is the case of the Swedish *focal accent*. But since there is usually no simple one-to-one mapping between form and function, as indicated in various places in this chapter (and exemplified for German in 3.2.3), Hirst (2005) argues that the two levels of form and function should be more carefully distinguished.

There are at least two alternative research paradigms for the investigation of intonation which make a rather explicit distinction between form and function. One perspective is to focus exclusively on form, explicitly ignoring functional aspects. The clearest example of this perspective is perhaps the research conducted in connection with the IPO model (e.g. Cohen and 't Hart 1967; 't Hart et al. 1990), originally developed for Dutch, but also applied to German (Adriaens 1991). One goal of this approach is to determine the inventory of distinctive intonational units of a language and the rules for their combination, in order to account for all possible intonation patterns of that language. The distinctiveness of patterns is not determined with reference to their function, but via their perceptual discrimination alone. The way in which this inventory is used, or which functions individual patterns may fulfil in communication, is regarded as a subsequent step of research (e.g. Terken 1985). This account can be referred to as a (strict) *form-first* approach.

A third perspective can be referred to as a *function-first* approach, which starts from previously selected functions and investigates how these functions are manifested in intonation. Typically, this approach implies an experimental setting where material is constructed in a way that would *elicit* the functions under investigation. A representative contemporary example of this approach are the studies by Yi Xu and colleagues in connection with the *Parallel Encoding and Target Approximation* (PENTA) model (e.g. Xu 1999; Xu and Wang 2001; Xu and Xu 2005; Xu 2005; Liu and Xu 2005; Prom-on et al. 2009; Xu 2009). While in most (phonological) approaches to intonation, the model categories are formal in nature (such as pitch accents and boundary tone), in the PENTA model, the central model categories are the *communicative functions*, each of which is associated with a unique *encoding scheme*. Surface $F_0$ contours are then derived from the *parallel encoding* (PEN) of functions, i.e. by a combination of encoding schemes, via a process of *target approximation* (TA).

The aspect of the PENTA model most relevant for our discussion is the idea of *parallel encoding*, which implies that prosody simultaneously fulfils multiple functions. One goal of the model is to isolate the contributions individual functions make to surface $F_0$. For that, in a typical study (e.g. Xu and Xu 2005; Liu and Xu 2005), a number of functions are selected and elicited simultaneously in a controlled manner. For instance, in Liu and Xu (2005) on Mandarin Chinese, the functions involved are lexical tone, focus, and interrogativity. Principally, this idea and method is similar to the one applied by Bruce (1977) or Bredvad-Jensen

(1981), who elicited lexical accent and focus (Bruce 1977; cf. 3.3.3) or focus, interrogativity, and different attitudes (Bredvad-Jensen 1981; cf. 3.3.4) in Swedish.

Elicitation studies following a function-first approach, even if combined with perception tests as in Liu and Xu (2005), can probably never suffice to reach a complete understanding of intonation, since they can hardly account for phenomena that would most likely occur in real human–human interaction only. Hence, it seems obvious that a function-first approach can only be a complement to form-inspired approaches, like investigations of spontaneous speech data. However, the experimental control it implies makes it suitable for certain research questions, such as the one formulated in this thesis, as outlined in 3.5.

# Chapter 3

# Swedish and German prosody

This chapter provides an overview of some aspects of German and Swedish prosody. First, it is briefly pointed out that German and Swedish share some basic prosodic features (lexical stress and quantity), although these features (mainly quantity) are not entirely comparable in the two languages (3.1).

Then, the actual point of departure for this thesis is treated in some detail, namely the basic tonal patterns of German and Swedish, as they are recognised by contemporary models. *Basic tonal patterns* is here meant to refer to such patterns that have received distinct phonological descriptions in the literature. In the case of German, all tonal aspects are related to intonation, while tonal aspects of Swedish include both intonation and word accents. For German, a variety of intonation models are available, which typically differ with respect to the number and types of intonation patterns accounted for. The goal of Section 3.2 is therefore to present a minimal set of basic intonation patterns of German, which would be agreed upon by most contemporary models. In 3.3, several studies on aspects of Swedish intonation and word accents that are relevant for the discussion of the inventory of basic tonal patterns are summarised, and the Lund model is chosen as a point of departure.

Section 3.4 attempts to unite the accounts of basic patterns of Swedish and German discussed in the previous sections by means of reference to the *nucleus* concept. Finally, the research questions are refined and an overview of the studies in this thesis is provided in 3.5.

## 3.1   Lexical stress and quantity

Both German and Swedish have lexical stress, i.e. one syllable in every simplex word is specified as primary stressed.[1] In both languages, stress placement follows certain rules (cf. Kohler 1995 for German; Bruce 1998 for Swedish; Zonneveld et al. 1999 for both West and North Germanic languages), but stress is still distinctive, i.e. there are minimal pairs which differ phonologically only in the placement of stress. Typical examples for German are verbal compounds such as ˈ*umfahren* 'to knock over' vs. *um*ˈ*fahren* 'to drive round'.[2] A similar example for Swedish is *köra* ˈ*på* 'to drive against (e.g. a wall)' vs. ˈ*köra på* 'to drive on (e.g. the road)'. A Swedish–German cognate minimal pair is ˈ*F/formel* (German/Swedish) 'formula' vs. *for*ˈ*mell* (German/ Swedish) 'formal'.

Swedish has a complementary quantity distinction which is connected to lexical stress (Elert 1964; Bruce 1998). A stressed syllable contains either a long vowel (plus an optional short consonant or cluster), i.e. V:(C)[3], or a short vowel plus a long consonant or cluster, i.e. VC:(C). The primary phonetic correlate of the phonological quantity distinction in Swedish is the relation of vowel and consonant *duration*, although a difference in *vowel quality* between short and long vowels can typically be observed, as well. However, the contribution of vowel quality is not equally large for all vowel pairs. While the quality difference is very pronounced in e.g. [ɑː] vs. [a], it is less salient in [iː] vs. [ɪ/i].[4]

In German, a quantity distinction can be observed only for vowels. Moreover, the distinction can be argued to be primarily one of vowel quality (Kohler 1995), and hence, the impressionistic labels *tense* (for long vowels) vs. *lax* (for short vowels) are also common for German. For most vowel pairs, the difference between *tense* and *lax* can be produced without a significant difference in duration, while the quality difference is large, such as in [i(ː)] vs. [ɪ]. The only vowel pair where practically no quality difference is observable and instead, quantity is primary, is [aː] vs. [a].

The quantity difference between German and Swedish can be illustrated by

---

[1]In compounds, there can be several stresses, where one is typically primary, the others secondary. In German, certain types of compounds can also maintain several equally prominent stresses (Kohler 1995).

[2]The IPA notation of primary stress (ˈ) is used here also in orthographic examples to mark the lexically stressed syllable.

[3]Possible syllable onsets are omitted, since they are not related to the quantity distinction. (C) marks optional final consonants or clusters, the latter of which are frequent in inflected forms, e.g. Swedish *lägst* 'lowest'.

[4]The transcription [ɪ] is not uncommon, but it is also typically used for the corresponding German or English 'short /i/' vowel, where the qualitative difference between [i] and [ɪ] is much larger than in Swedish. The transcription [ɪ/i] has been chosen to stress that the Swedish 'short /i/' is often actually close to [i].

the minimal pair *Polen* 'Poland' vs. (German/Swedish) *P/pollen* 'pollen', given in (1) in a broad phonetic transcription:

(1)    German:    [ˈpʰo(ː)lən] vs. [ˈpʰɔlən]
         Swedish:    [ˈpʰoːlɛn] vs. [ˈpʰɔlːɛn]

## 3.2    German intonation

### 3.2.1    Models of German intonation

Numerous descriptions or models of German intonation have been formulated within a variety of schools. Examples are von Essen (1964), Isačenko and Schädlich (1970), Bannert (1984), Kohler (1987, 1991a, 1997), Wunderlich (1988), Altmann et al. (1989), Adriaens (1991), Uhmann (1991), Möbius (1993), Féry (1993), Grabe (1998), and Grice and Baumann (2002). While most accounts of German intonation differ crucially in one or the other conceptual or methodological respect, as well as in the exact number of categories in the postulated inventory of German intonation patterns, there is nevertheless a wide consensus about some basic issues. It is, for example, possible to pinpoint some general patterns or pattern types (such as a *simple rise* or a *fall-rise*) whose existence all contemporary approaches would agree upon. The purpose of Section 3.2.3 is to present and describe an inventory of such basic intonation patterns of German as well as a selection of suggestions concerning the *functions* or *meanings* of these patterns.

   Since only the very general common ground of the available models is relevant for this thesis, none of the individual accounts is discussed in detail here. However, in order to describe the basic inventory of German, a transcription system is required. Regarding Swedish, this thesis is based on the Lund model (cf. 3.3.3), which is formulated in AM terms. Thus, for the sake of comparability, an AM-based transcription system should be chosen even for German. Two obvious candidates among contemporary accounts of German intonation are Féry's (1993) model and GToBI (German Tones and Break Indices, Grice and Baumann 2002). Féry's (1993) tonal representations are intended to represent assumed underlying, rather than surface patterns.[5] That is, Féry (1993) offers primarily a phonological model rather than a transcription system. GToBI is, although based on phonological theory, actually a broad-phonetic transcription system. The Lund model label inventory is far less detailed than GToBI, but still rather surface-oriented. Therefore, the GToBI system, with a slight simplification, is chosen for the purpose of

---

   [5]For instance, a low pitch on the stressed syllable in connection with an 'early peak accent' (cf. 3.2.3), transcribed as H+L*, i.e. with L* (low tone) in GToBI, is represented as H+H*L, i.e. with an (assumed underlying) H* (high tone) by Féry (1993). Another example is that phrase boundaries do not need to be tonally specified in Féry's (1993) model.

describing German intonation patterns. It is presented in 3.2.2. However, for the presentation of the functional characteristics of the basic patterns of German in 3.2.3, reference is also made to further accounts of German intonation.

## 3.2.2   Transcription of German intonation

GToBI (Grice and Baumann 2002; Grice et al. 2005) is the German adaptation of the ToBI system (e.g. Beckman et al. 2005), which was originally designed for American English intonation.

According to GToBI, each *intonation phrase* (IP) comprises at least one *intermediate phrase* (ip); an IP containing more than one ip thus exhibits minor phrase boundaries between the ips. Furthermore, each ip contains at least one *pitch accent*[6], and exactly one *phrase (edge) tone*, which accounts for the tonal pattern between the final accent and the final edge of the ip. This tone is either H or L (or !H, i.e. downstepped H), and its functions as a phrase tone is indicated by a '-' as a diacritic (H-, L-, !H-). Finally, each IP also contains an H or an L as a final *boundary tone*, which is symbolised as H% or L%. That is, each IP ends with a combination of a phrase tone (H-, L-) and a boundary tone (H%, L%). Where the two tones have the same value, the transcription is simplified, e.g. L-% instead of L-L%. Four basic final boundary patterns are assumed for an IP in GToBI: L-%, H-%, L-H%, and H-ˆH%, where ˆH denotes an upstepped H tone.

This assumed IP structure implies some phonotactic rules for the elements in Table 1.1 (Chapter 1). One of them has already been incorporated in Table 1.1, namely the obligatory phrase tone preceding a boundary tone. Furthermore, a phrase or boundary tone (i.e. H-, L- or any of the four complex boundary tones) must always be preceded by an accent (which can be several syllables or words away from the final boundary). The last accent in an IP is referred to as the *nuclear accent* (cf. 2.1.2) of the IP, and the entire complex of nuclear accent and boundary tone pattern is referred to as the *nuclear pattern* in this thesis.

A pitch accent in GToBI consists either of a single tone or a combination of two tones, connected in notation by a '+'. The tone that associates with the tone-bearing unit, i.e. the stressed syllable in German, is *starred* (marked by '*'); the other tone, in case there are two, *leads* or *trails* the starred tone. GToBI assumes six German accents, cf. Table 1.1. For the description of the basic pattern inventory in 3.2.3, four of these will be used distinctively, namely H*, L*, H+L*, and L*+H.

This thesis is not concerned with the issue of different levels of phrasing. Therefore, a simplification of the GToBI system concerning phrase boundaries

---

[6]"Exceptions to this rule are the so-called 'intonational tags', which can be regarded as enclitic tone units without pitch accents." (Grice et al. 2005)

is introduced, by omitting the '-' diacritic in the GToBI transcriptions.[7] As a further simplification, only one of the three high-ending boundary tone combinations (LH%, H-ˆH%, H-%, cf. above) is regarded necessary for the description of basic patterns in 3.2.3. Hence, the two final phrase boundary tones used to describe the basic patterns of German in this thesis are L% (low) and LH% (low-high). It should be stressed that these simplifications are not introduced in order to suggest any modifications of the GToBI system, but only to provide a simple inventory of symbols, which are comparable to the symbols used in the Lund model (cf. 3.3.3).

### 3.2.3 Basic intonation patterns of German and their communicative functions

Both the *paradigmatic* and the *syntagmatic* dimension of intonation (cf. 2.2.1) have played an important role in German intonation research, although many studies have focused primarily on one of the dimensions. For instance, Uhmann (1991) concentrates on the *syntagmatic* dimension (primarily *placement* of accents for focus signalling), while the Kiel Intonation Model (Kohler 1991a) is more concerned with the *paradigmatic* dimension (choice of accent *type* for the signalling of *relational* (cf. 2.2.3), including attitudinal meanings). The introduction to paradigmatic and syntagmatic intonational contrasts in 2.2.1 to 2.2.3, although mainly based on English, is basically valid also for German. This section concentrates mainly on the paradigmatic dimension, since the point of departure for this thesis is the striking difference in the inventory of paradigmatically contrasting patterns in German and Swedish as assumed by contemporary models (cf. Table 1.1 in Chapter 1).

The full repertoire of paradigmatic accent contrasts is sometimes regarded to be restricted to the nuclear position (e.g. Féry 1993). Therefore, this section focusses on *nuclear patterns* (nuclear accent plus final boundary pattern), which are exemplified by means of short, single-accent utterances, where the nuclear pattern basically represents the intonation of the entire utterance. A total of *five nuclear patterns* of German are assumed here to constitute a *minimal basic inventory* of German, in the sense that hardly any contemporary account of German intonation would disagree about their existence.[8] This basic inventory is in fact almost equivalent to the inventory proposed by Féry (1993).[9] An additional, very typical

---

[7]As seen in 3.3.3 below, this simplifiation also avoids an ambiguity when transcriptions for German and Swedish are compared, since in the Swedish context, the '-' sign is used with a different interpretation, namely to mark an *accent*, rather than a *boundary tone* as in GToBI.

[8]The inventory is minimal in the sense that some models of German intonation (e.g. Grice et al. 2005; Kohler 1991a) assume further pattern types (or modifications of them) not taken into account here, which, however, differ between the models.

[9]Only the stylised calling contour is disregarded here.

Table 3.1: Six basic intonation patterns of German, represented by a simplified GToBI transcription.

|   | Pattern | Transcription | Usage (selection) |
|---|---------|---------------|-------------------|
| 1. | *early fall* | H+L* L% | established, confirmed, unchangeable facts; summarising, closing an argument; accessible discourse referents |
| 2. | *medial fall* | H* L% | opening an argument; new, negotiable information; *default* in declaratives, *wh*-questions, wishes, imperatives |
| 3. | *late fall* | L*+H L% | opening an argument; new, unexpected information; (emphatic) surprise; indignation; implying 'of course' |
| 4. | *fall-rise* | H* LH% | *yes/no*-questions;  polite offer;  requests; friendly warning |
| 5. | *simple rise* | L* LH% | questions (echo; *yes/no*-; friendly *wh*-); non-finality; uncertainty |
| 6. | *hat pattern* | L*+H... H+L* L% | contrastive topic ...  focus; semantic cohesion; *default* concatenation of certain accents |

pattern of German is the *hat pattern*, which results from the combination of at least (and usually exactly) two pitch accents, a *pre-nuclear* and the *nuclear* one. These six basic patterns of German are represented in simplified GToBI notation (cf. 3.2.2) in Table 3.1. The table includes a selection of suggestions that have been made concerning the function, meaning, or usage of the individual patterns. This selection is primarily based on Kohler (1991b, 2006a), Féry (1993), Grice et al. (2005), Baumann (2006), Niebuhr (2007b), and Ambrazaitis and Niebuhr (2008).

### Realisation of the five basic nuclear patterns

The approximate, typical realisations of the basic patterns are illustrated in what follows mainly for short utterances like *November* [noˈvɛmbɐ], where at least one unaccented syllable is preceding and following the stressed one. A basic distinction can be made between patterns with a final rise in $F_0$ (patterns ending in LH%) and a final fall (patterns ending in L%). A *simple rise* (L* LH%) in the utterance *November* would typically be realised with low $F_0$ on the first two syllables, and

a rise in the last syllable.[10] In a *fall-rise* (H* LH%), the first syllable is low, followed by a peak in the stressed syllable and a falling-rising pattern in the final syllable.

As indicated by the labels *early*, *medial*, and *late*, the three-fold distinction of falling patterns has been described in terms of the *timing* of a pitch peak.[11] The first comprehensive account of this distinction has been presented by Kohler and colleagues (Kohler 1987, 1991b, 1991a; Kohler and Gartenberg 1991; Gartenberg and Panzlaff-Reuter 1991; Hertrich 1991a, 1991b), and further developed by e.g. Dombrowski (2003), Kohler (2005c), and Niebuhr (2007b). In an *early fall* (H+L*), $F_0$ is relatively high in the pre-stress syllable in an utterance like *November*, and low or falling in the stressed syllable, which (perceptually) results in a low pitch.[12] In the final syllable, $F_0$ (and pitch) usually continues to fall slightly.

In a *medial fall* (H* L%), $F_0$ is high or rising in the stressed syllable, usually resulting in high pitch, while pitch is relatively low in the pre- and the post-stress syllable. That is, the $F_0$ peak is located later (i.e. in the stressed syllable) in H* than in H+L* L% (cf. above). Finally, in a *late fall* (L*+H L%), the peak is shifted yet further into the stressed syllable or into the post-stress syllable. The latter case is likely only when further unaccented syllables follow the stress, like in Niebuhr's (2007b) example *eine 'Malerin* 'a (female) painter'. In an utterance like *November*, where only one syllable follows the stress, the final syllable also has to accommodate the utterance-final fall, and hence, the peak of L*+H will usually be realised already in the transition from the stressed to the post-stress syllable.

Hence, the phonetic (both acoustic and perceptual) distinction between H* and L*+H is often less clear than between H* and H+L*.[13] In general, and specifically in a context like *November*, the distinction between *medial* (H*) and *late* (L*+H) is not only a matter of the timing of the $F_0$ maximum (which can be located near

---

[10]For the purpose of presenting only the basic pattern types, no distinction is made here between different, well-attested versions of the rising pattern, as described in e.g. Kohler (1991a) or Grice et al. (2005). The rise most relevant for the present discussion would actually be represented as L* H-ˆH% in GToBI.

[11]In the literature on German, the three basic patterns referred to here as (early, medial, or late) *fall* are typically referred to as (early, medial, or late) *peak*. The latter notion focusses primarily on the type of accent, which might be nuclear or pre-nuclear. The notion of (early, medial, or late) *fall* is used here to refer to the entire *nuclear pattern*, including the final fall (cf. Table 3.1).

[12]Another variant of an early fall has been suggested (Grice et al. 2005), where pitch is relatively high also in the stressed syllable, but still lower that in the pre-stress, represented H+!H* (downstepped H) instead of H+L*. For discussions and studies on the distinction, see, e.g. Rathcke and Harrington (2006) and Grice et al. (2009).

[13]Categorical perception (CP) has been found only in the transition from H+L* to H*, not from H* to L+H* (Kohler 1987, 1991b). However, the relevance of CP in intonation for the distinction between phonological categories has been discussed and rejected by Niebuhr and Kohler (2004), Niebuhr (2007a).

the offset of the stressed syllable in both cases), but also a matter of the shape and height of the $F_0$ peak. For instance, in H* the $F_0$ peak can be broader, resulting in high pitch throughout the vowel, while in L*+H there is a more pronounced rise in the vowel, often enhanced by means of a higher $F_0$ maximum.[14] For a comprehensive account of the perception of the *early*, the *medial*, and the *late fall* and their phonetic correlates (including peak shape and height, timing, and intensity) see Niebuhr (2007b).

**Functions and meanings of the five basic nuclear patterns**

While there is a broad consensus about the existence of the six basic intonation patterns displayed in Table 3.1, it is less clear what the communicative functions of these patterns are. The usage of an intonation pattern can have a variety of different pragmatic interpretations, and hence, the meaning differences associated with paradigmatic intonational contrasts are often elusive (cf. 2.2.3). Several factors, for instance discourse structure and speaker attitude, are relevant. For some patterns, attempts at pinpointing general, underlying semantic features have been made, in a similar manner as those done for English (cf. 2.2.3; Gussenhoven 1984; Pierrehumbert and Hirschberg 1990). The presentation here summarises some proposals made in the literature.

The basic distinction between a final rise (...LH%) and a final fall (...L%) has, for instance, been related to the signalling of sentence mode. The text-book statement is that falling intonation (usually in connection with the default accent H*) occurs in declarative sentences and *wh*-questions, but also in some wishes and imperatives (Féry 1993), while rising intonation occurs in *yes/no*-questions. However, as mentioned in 2.4, this distribution of rises and falls is only a tendency. It interacts with the signalling of the speaker's attitude towards the listener and the communicative situation, a rise signalling *openness*, a fall *categoricalness* (Kohler 2005b). A final rise has also been associated with functional notions such as *non-finality* or *incompleteness* (e.g. Féry 1993) in the sense that the speaker expects some reaction from the listener. Certainly, *non-finality* can also be signalled in connection with a *fall* (if all questions are assumed to imply non-finality in some sense), while *rises* typically do not signal *finality* in German (e.g. Ambrazaitis 2005).

The functional distinction between the *simple rise* (L* LH%) and the *fall-rise* (H* LH%) has not been treated systematically yet. In general, the *simple rise* seems to be more common in questions than the *fall-rise*, while the *fall-rise* has

---

[14]The distinction between H* and L*+H is further complicated by the fact that further accent types have been suggested which are related both to H* and to L+H*, namely, first, an L+H*, where the rise is more pronounced than in H* (Grice et al. 2005), and second, a *late-medial* peak, which is assumed to constitute a timing position between medial and late (Kohler 2005c).

been associated with various pragmatic meanings. For instance, it may be used in (friendly) warnings (Kohler 2006a), typically in child-directed speech, e.g. *Du!* 'you' in a context like *Du! Jetzt ist Schluss!* 'Will you stop that!'. However, it is also common in requests, or in addressing a person in order to make a request, not seldom uttered by children to their parents, as in *Papa?* 'Daddy?' (cf. the acknowledgments in Kügler 2007). The fall-rise can also be used in questions (e.g. Ambrazaitis 2005) or polite offers (Grice et al. 2005).

The three-fold contrast of falling patterns (*early*, *medial*, *late*), or subsets of it, have received much attention from different perspectives. According to Kohler (1991b), the early fall signals an "established fact and the summing up of an argument", the medial fall signals "a new fact and opens up a new argument", and the late fall signals "emphatic surprise". In his classical perception experiments, Kohler (1987, 1991b) constructed the contexts displayed in (2) in order to elicit *matching* responses between a pattern – *early*, *medial*, *late* – and the hypothesised function, represented by a, b, and c, respectively, in (2), here illustrated for the test phrase *Sie hat ja gelogen.* 'She's been lying', indicated by . . . in (2); only the English translation is shown.

(2)  a.  'Once a liar, always a liar. This also applies to Anne. . . . '
     b.  'Now I understand. . . . '
     c.  'Oh! . . . '

Baumann and colleagues (Baumann and Hadelich 2003; Baumann and Grice 2004, 2006; Baumann 2006) investigated the distinction between H+L* and H* from the perspective of *givenness* and *degrees* of givenness in the sense of Chafe (1994), where three basic cognitive states of discourse referents are assumed: *given*, *accessible*, and *new*. In a priming study, Baumann and Hadelich (2003) found that an H* is preferred for *new* information, and deaccentuation is preferred for *given* information, while no specific preference could be established for *accessible* information. In a follow-up study, Baumann and Grice (2006) elaborated on different types of *accessibility*. For instance, a discourse referent can be accessible due to previous, but not immediately preceding, mentioning (*textually displaced*), due to lexical-semantic relations (e.g. *synonymy* or *hyponymy*), or due to a *scenario* relation (e.g. bus – driver). The results were, in summary, that an H+L* was preferred over an H* for all types of accessibility, while deaccentuation was still preferred over H+L* in five of the eight types tested. However, in two types of accessibility (*whole–part* relation and *scenario*), an H+L* was preferred over deaccentuation. To summarise both studies, an H* is most adequate for *new* information, deaccentuation for *given* information, but an H+L* is most adequate for some types of *accessible* information.

Baumann's (2006) studies show that the cognitive status of discourse refer-

ents is *one* factor that *can* play a role in the choice of accent. However, it is obvious and also admitted by Baumann (2006) himself, that further factors are important. After all, the speaker still has the choice to mark information as established, new, or surprising (cf. Kohler 1991b), i.e. to choose any of the falling patterns in order to express his/her *attitude* towards a certain discourse referent. As Baumann (2006) puts it, "if a speaker wishes to present a constituent as particularly newsworthy, s/he can highlight this constituent irrespective of its activation status [i.e. degree of givenness, ga]. [...] Thus, the focus–background structure of an utterance may override the level of activation by superimposing a pragmatic level, which indicates relations between discourse referents and propositions, on the more 'objective' level of the referents' mental states".[15]

The conceptual independence of the choice of accent from the status of a discourse referent in terms of givenness can be illustrated by means of (potential) real-life examples. An example discussed by Niebuhr (2007b) is the usage of the H+L* by train staff in the announcement of a revised arrival time. Here, the H+L* does not imply that the delay has been mentioned before or is accessible in some way, but rather that the information represents an unchangeable fact. The aspect of routine might also play a role in this example.

In a similar manner as Gussenhoven (1984) has suggested underlying *meanings* for three basic patterns of English and Dutch (corresponding to the *medial fall*, the *simple rise*, and the *fall-rise* in Table 3.1), Niebuhr (2007b) has proposed such underlying meanings, or *generic characteristics* ("generische Characteristika") for the three falling patterns of German, namely GIVEN for H+L*, NEW for H*, and UNEXPECTED for L*+H. Niebuhr (2007b) argues that these basic meanings can receive various pragmatic interpretations, covering both Baumann's (2006) findings and the traditional descriptions (Kohler 1991b). In the same vein, but using partly different terms, Kohler (2006a) proposes to summarise the meanings of the three patterns *early*, *medial*, and *late*, as *finality*, *openness*, and *unexpectedness*.

### Forms and functions of the *hat pattern*

The *hat pattern* (here understood in the sense of a *flat hat* in 't Hart et al. 1990) contains at least two accents, like in *PEter kommt im NoVEMber* 'PEter is coming in NoVEMber' (capitals indicate accents). Thus it does not represent an additional basic *nuclear pattern* like the other five basic patterns discussed above, but a structure comprising both a *pre-nuclear* and a *nuclear* accent.

---

[15]This insight is similar to Bolinger's (1972) much cited paper title that "Accent is predictable (if you're a mind-reader)". However, Bolinger (1972) was concerned with the *placement* of accents, while Baumann (2006) quotation is also concerned with the *choice* of accent *type*.

The constituting feature of a *hat pattern* is a high pitch plateau connecting the two accents. A hat pattern can contrast against a dipped concatenation, i.e. a sequence of two accents with a clearly pronounced pitch peak each. In most AM models, a dipped sequence and a hat result either from different accent combinations or they reflect a phrasing difference. For instance, while the sequence L*+H H+L* would necessarily result in a hat (Féry 1993), L*+H L*+H would imply a dip. Another accent sequence (H* H*) can be produced with an intervening phrase boundary (H* L% H*), which results in a dip, while it usually surfaces as a hat, when produced within the same phrase, which is, however, depending on the temporal distance between the accents (Féry 1993). That is, if the distance becomes large enough, the concatenation can be dipped (cf. the 'sagging transition' in Pierrehumbert 1980 for American English).[16] An alternative approach (e.g. Kohler 1991a; Peters et al. 2005; Ambrazaitis and Niebuhr 2008) is to regard the concatenation (hat vs. dip) as a feature independent of the accent combination or of intervening phrase boundaries.

A general function that has typically been attributed to the hat pattern is the signalling of *cohesion* (e.g. Féry 1993; Peters et al. 2005; Ambrazaitis and Niebuhr 2008). The hat pattern has also been associated with *broad focus* or regarded as the more *neutral* pattern, while a dipped sequence signals a *double (narrow) focus* or more *emphasis* (Peters et al. 2005).

In general, different combinations of accent types are possible in a hat pattern in German (e.g. Kohler 1991a; Hertrich 1991b; Féry 1993; Peters et al. 2005). This thesis treats the hat pattern only as a secondary object of study (cf. 6.1.2), and focusses on the combination of accents displayed in Table 3.1: L*+H in the first position, H+L* in the second position.

While the general function of cohesion signalling also applies to this particular hat pattern (L*+H H+L*), it has been reported (e.g. Féry 1993; Jacobs 2001) to have a more specific function related to *information structure*, namely the signalling of a CTOP+FOC structure (cf. 2.3.2): the pre-nuclear rise in the hat pattern (L*+H) marks a *contrastive topic* (CTOP), while the nuclear early fall (H+L*) signals *focus* (FOC).[17]

Although *focus* signalling has not been treated explicitly in the discussion of

---

[16]However, in a spontaneous speech corpus, Peters et al. (2005) found no salient difference in mean duration between hat patterns and dipped sequences, and the standard deviations were large for both samples. Furthermore, the study by Niebuhr (2005) suggests that a dip/hat contrast can (acoustically and perceptually) be maintained even when the two accents are placed on short vowels in syllables immediately following each other.

[17]Féry (1993) uses the label H*L for the nuclear accent in the hat, suggesting a high underlying accent tone (H* in our notation). Féry (1993) does not distinguish phonologically between a potential distinction of accent type in the second position in a *hat*, but her H*L accent can be interpreted both as a *medial* and an *early* fall (cf. Footnote 5).

the basic nuclear patterns above, it can be inferred that the typical accent used for signalling *focus*, at least when the focus renders *new information* (cf. 2.3.1), is an H* or an L*+H, rather than an H+L* (cf. Table 3.1). However, in the hat pattern discussed here, an H+L* can be used for signalling *focus*, even if the focussed information is *new* (as in example (12) in 2.3.2). A perception-based explanation is proposed by Kohler (1991a), who argues that an H* in final position in a hat actually represents an "attenuated *early peak*" rather than a *medial peak* proper, since it is preceded by the high plateau. This "attenuated *early peak*" is perceptually less prominent than a (regular) early peak (H+L*) on the one hand, or as a medial peak (H*) on the other hand. Therefore, if a hat pattern is used in a context which demands the second accent to be more prominent than the first, then H+L* is to be preferred over an H* in the second position.

## 3.3    Swedish intonation and word accents

This section introduces the basic repertoire of tonal patterns of Swedish as assumed in the Lund model (3.3.3), which serves as a point of departure for the studies in this thesis. In addition, some further studies are summarised, which are not directly related to the development of the Lund model, but relevant for the discussion of the basic tonal pattern inventory of Swedish (3.3.4). The word accent distinction is also treated in some detail (3.3.1), since it has traditionally been one important driving force in Swedish prosody research and is obviously highly relevant for the modelling of basic tonal patterns in Swedish.

### 3.3.1    The Swedish word accent distinction

The majority of Swedish dialects, including the Standard variety, have a so-called *word accent*[18] distinction between *accent I* and *accent II* (or *acute* and *grave* accent). This distinction is present also in Norwegian, and a related phenomenon, the *stød* (as opposed to *no stød*), occurs in Danish. The word accent contrast of Swedish and Norwegian is primarily tonal. Much effort has been spent on the investigation of diachronic aspects of the word accent distinction (e.g. Riad 1998a, 2005 and references therein), which are, however, not dealt with here. The main focus of the present introduction is the phonological representation and the phonetic manifestation of the word accents, as well as the interplay of word accent realisation and intonation. Some notes on the function of the word accent distinction are also made. The discussion deals primarily with $F_0$ as the main acoustic

---

[18]Word accents are also referred to as lexical (pitch/tone) accents, tonal accents, word tones, or related notions in the literature.

correlate of the word accents (e.g. Malmberg 1959), and with the Standard variety of Swedish, but where relevant, reference is made to other dialects. For more general introductions to the Scandinavian word accents see, for instance, Gårding (1977) or Bruce and Hermans (1999).

**Tonal patterns in citation forms**

A word with accent I is typically spoken with a different $F_0$ pattern than a word with accent II (Meyer 1937), and this difference in $F_0$ is the most important cue from a perception and recognition point of view (Malmberg 1966; House and Bruce 1990), even if there are further secondary phonetic correlates such as duration or intensity (Malmberg 1966).[19] The prosodic domain of a *'word' accent* is, in general, not a *word* in the sense of a lexical or a grammatical unit (Haugen and Joos 1972), but rather a *prosodic word*, which usually corresponds to a content word and adjacent unaccented function words (cf. 2.1.2).

Meyer (1937) conducted the first large-scale systematic instrumental investigation of the word accent contrast in a variety of Swedish dialects. His results are the basis for what is often referred to as the 'traditional' understanding of the word accent contrast. According to this traditional view, when a word is spoken in isolation (i.e. as a one-word utterance), the tonal pattern of Standard Swedish accent I is a rise through the stressed syllable, followed by a fall in the post-stress syllable.The accent II pattern, however, comprises a fall in the stressed syllable, followed by a rise and a second fall in the post-stress syllable (in simplex words) or in the secondary stress (in compounds). Accent I has also been referred to as 'single-peaked' and accent II as 'double-peaked'. Dialects differ crucially in their realisation of the word accents. For instance, in many dialects, both accents are 'single-peaked', and the difference between accent I and II is a matter of timing or shape of the contour (Meyer 1937; Malmberg 1953; Bruce and Gårding 1978).

**Functional aspects**

Word accent is a distinctive feature in Swedish phonology, since it can distinguish between words, such as *regel* (accent I) 'rule' and *regel* (accent II) 'bolt'. However, distinctiveness is not the primary function of the word accents. There are only about 350 minimal pairs in Swedish (Elert 1972) and in many cases, one member of the pair is rather unusual. Moreover, word accent is largely determined by the morphological and phonological structure of the word. Thus, in many cases word accent is only distinctive in inflected forms, such as in *anden*

---

[19]In the beginning of the 20th century, it was still a matter of debate whether the word accent distinction was primarily a pitch contrast or an intensity contrast, i.e. whether the word accents were tonal (or melodic) or dynamic (cf. the introduction in Meyer 1937).

(accent I, definite singular of *and*) 'the duck' vs. *anden* (accent II, definite singular of *ande*) 'the spirit'. Often, different parts of speech are involved in a minimal pair, which usually cannot contrast paradigmatically in natural discourse, such as *buren* (accent I, definite singular of *bur*) 'the cage' vs. *buren* (accent II, perfect participle of *bära*) 'carried'. Examples for rules of word accent assignment are, for instance, that in most Swedish dialects, all words with final stress, including all monosyllabic words, receive accent I[20], and in Standard Swedish, basically all compounds receive accent II. For more detailed rules, also with respect to inflectional morphology, see e.g. Elert (1972) or Gårding (1977).

Thus, the primary function of the word accents is related to morphology or morpho-syntax. It has been proposed that accent II has primarily a *connective function* (Malmberg 1959; Elert 1970), which aids the listener in processing morphological structure. That is, perceiving an accent II pattern on the stressed syllable indicates that at least one further syllable will follow which belongs to the same word as the stressed one. The interaction of the tonal pattern of the stressed syllable and a suffix in disyllabic words has recently been investigated in a neurolinguistic study (Roll 2009) using Event-Related Potentials (ERP), as discussed in more detail below.

Moreover, word accent plays an important role in the signalling of regional origin. Second language learners who – consciously or unconsciously – choose to ignore the word accent contrast, are still perfectly understood, at least in case their pronunciation is acceptable otherwise, but the lack of the word accents contributes to their foreign accent. Furthermore, the phonetic realisation of the word accents differs crucially from dialect to dialect, and finally, there are also Swedish dialects which lack the word accent distinction (e.g. Finland Swedish dialects).

**Separating word accents from intonation**

Meyer (1937) was aware of the fact that some aspects of the patterns found in his data (short utterances, mostly citation forms) must be due to the *sentence intonation contour* ("Satztonkurve"). In particular, he assumed that sentence intonation involves an initial rise and a final fall, and hence the tonal pattern of a word will be more affected by sentence intonation when the word is placed in sentence initial or final, than in sentence medial position. But Meyer (1937) concluded that it would hardly be possible to decompose the observed tonal pattern in order to isolate the influence of the word accents on the one hand and sentence intonation

---

[20]That is, there is actually no word accent distinction in words with final stress, and hence, some authors have suggested that such words are neither associated with accent I nor with accent II (e.g. Malmberg 1966).

on the other hand.[21]

Jassem (1962), in a study on a short read text, followed by a small-scale perception experiment, found that the word accents were realised with highly variable patterns. One typical pattern was the one that corresponds to the traditional description (cf. Meyer 1937), but another frequent pattern was a *fall*, which for accent I always implied a high pitch level in the pre-stress syllable. In the case of such a falling pattern, accent I and accent II could still be distinguished by means of the tonal relation between the (high) pre-stress and the stressed syllable, the stressed one being lower in accent I than in accent II. According to Jassem (1962), this variation of word accent patterns may be somehow related to position in the utterance, but no further clarification was suggested.

An attempt to isolate the influence of word accent and intonation on the melodic pattern of an utterance was presented by Öhman (1967, 1968) who proposed a functional larynx control model where $F_0$ contours are modelled indirectly by simulating the underlying production process. Word accent and sentence intonation are modelled as separate, relatively simple step functions representing neural commands. Sentence intonation is modelled as a positive pulse, while word accent is modelled as a negative pulse. The intonation pulse by itself results in a rising-falling $F_0$ movement which is roughly similar to the traditional description of accent I in a citation form. The negative word accent pulse has the effect of eliminating portions of the high $F_0$ stretch resulting from the intonation pulse. The most crucial difference between accent I and II is a matter of the *timing* of the negative pulse. The effect of the *early* pulse used for accent I is that the onset of the $F_0$ rise predicted by the intonation pulse is lowered, resulting in a slight delay of the $F_0$ rise, corresponding better to the empirical accent I pattern. For accent II, the negative pulse is timed *later* and hence introduces an $F_0$ dip in the mid part of the sentence intonation peak, while $F_0$ before and after this dip remains high, resulting in the characteristic two-peaked contour for accent II. Öhman's (1968) model is further discussed below.

In a comparison of Meyer's (1937) data for different dialects, Gårding (1970) observed that the last portion of an accent II contour in a disyllabic word (i.e. the second peak in case there are two peaks in the given dialect) always corresponds to the first portion of an accent I contour (i.e. the single peak). Gårding (1970) concluded that only the first portion of accent II (i.e. the first peak in case there are two) is a 'real word tone', while the accent I contour, as well as the final part of accent II, represents sentence intonation.

Bruce (1977) elicited a corpus of read test sentences where word accent and *focus location* were experimentally controlled, the latter by means of context ques-

---

[21]"Eine empirische Intonationskurve in die beiden Komponenten, aus denen sie zusammengesetzt ist, mit Sicherheit zu zerlegen, ist nun freilich kaum möglich." (Meyer 1937)

tions. The second tonal peak in accent II, or the only one in accent I, respectively, occurred only if the word was in *focus*. Hence, he could confirm Gårding's (1970) identification of the second peak as a *sentence accent* (or *focal accent*), and analysed it as a high tonal target (H). The pitch fall that would follow his H in a citation form could be identified as a sentence final boundary signal, and consequently analysed as a low tonal target (L).

The sentence accent also provided a key to a novel understanding of the word accents, since it implies that in sentences consisting of more than one prosodic word there are both *focal* and *non-focal* realisations of the word accents. Bruce (1977) showed that even when a word was not in focus, i.e. when the focal accent was placed on a previous or a following word in the sentence, the non-focal word was still associated with a specific $F_0$ pattern. For accent I words, there is typically a falling $F_0$ movement from the pre-stress syllable to the stressed syllable, rendering a low-pitched stressed syllable. In accent II, there is also a fall, which, however, occurs later than in accent I, rendering a high-pitched stressed syllable onset and a fall through the vowel. This finding regarding pre- and post-focal realisations of accent I and accent II might explain the variation of word accent patterns found by Jassem (1962), cf. above, and led to the analysis of the word accent distinction in terms of a *timing* contrast of a pitch fall.

Thus, Bruce (1977) demonstrated that the tonal pattern of a Stockholm Swedish sentence is made up of three components: the word accents, the sentence accent, and the final boundary. This implies a revision of the traditional view in which accent II is 'double-peaked' and accent I 'single-peaked', as usually observed in citation forms. That is, according to Bruce (1977), both accent I and II are single-peaked when non-focal, but double-peaked when focal, at least at an underlying level (cf. next section).

Bruce's (1977) analysis is to some degree in line with Öhman's (1968) model, since both assume a sentence accent, and both account for the word accent distinction in terms of timing. However, there are crucial differences. First, by varying focus in the test material, Bruce (1977) could more accurately identify the sentence accent (the second peak in case of accent II), which falsifies Öhman's (1968) assumption that the sentence intonation gives rise to both peaks in the case of accent II. Second, Bruce (1977) accounts for non-focal accents. Third, as a consequence of the first two points, Bruce (1977) arrives at different tonal representations for the word accents (tonal falls) than Öhman (1968) (whose negative pulses can be interpreted as simple low tones).

**Phonetics, phonological representation, and lexical specification**

There have been and still are controversies concerning the issue of 'markedness' of the word accents. Opposite proposals may, in part, be due to the fact that

markedness is discussed in the context of two phenomena, which are separated by some scholars, but regarded as intimately connected by others, as discussed by Riad (2009), namely the *phonological representation* on the one hand, and the *lexical specification* of the word accents on the other hand. The issue with phonological representation is whether both accents, or only one of them, has a specific tonal correlate and can hence be positively specified in phonological terms. The issue with lexical specification is which words receive their accent by (post-lexical) default rules, and which words (or morphemes) are specified in the lexicon for either accent I or II. Riad (2009) comments on the fact that these two issues sometimes are treated separately and argues that it is only meaningful to speak of markedness (which he relates to lexical specification) in case markedness is reflected in representation. We are here not interested in the theoretical notion of markedness as such and will hence not discuss it in full detail. However, the results of different studies in this thesis may have an impact both on the discussion of the *lexical specification* (which is related to the perception or processing of the word accents) and the *phonological representation* of the word accents (which is related to their phonetic manifestation). Hence, some of the main arguments for different accounts are summarised in this section.

The most common view is to regard the word accent distinction as *privative*, where only accent II is assumed to be tonally specified, or to represent a real tone accent. A classical argument is that accent I has a less complex tonal pattern, representing "merely stress" (e.g. Witting 1977) like in other Germanic languages like German and English. That is, accent I reflects "the usual sentence intonation without any modifications" (Elert 2000, our translation). Also, when Swedes speak another language, they only use their accent I pattern. Accent I is hence a default pattern, which is also assigned to loan words (e.g. Witting 1977). This privative view is advocated by, e.g. Rischel (1963), Haugen (1967), Gårding (1970), Elert (1972), Witting (1977), Engstrand (1995, 1997), Riad (1998b, 2006).

An alternative view is the *equipollent* treatment of the word accents, where both accent I and II have a tonal specification (alternatively a model representation, e.g. in terms of *accent commands*). Two prominent, although different, examples have already been discussed in the previous section (Öhman 1968; Bruce 1977). In both accounts, accent I and accent II are assumed to be associated with the same type of tonal pattern, the distinction being a matter of timing.[22] But it must be stressed that this equipollent account is primarily concerned with the phonetic realisation and the phonological representation of the word accents, while it still can acknowledge an *asymmetric relationship* between the accents. As most

---

[22]In the model by Fujisaki et al. (1993), which is based on Öhman (1968), the distinction is rather a matter of the structure of the accent command; e.g. in citation forms, it can take the form of a sequence of two pulses: positive–negative for accent I, and negative–positive, for accent II.

clearly expressed in Gussenhoven and Bruce (1999), Bruce regards accent I as a default in the sense summarised above, while still assuming an equipollent phonological representation. This discussion is continued below.

In a study using Electromyography (EMG), Öhman et al. (1967) measured the activation of the *vocalis* and the *cricothyroid* muscles during the production of the word accents in order to provide empirical evidence for the modelling of the word accents as negative pulses as suggested by Öhman (1967).[23] However, the results of the EMG study were inconclusive. In fact, Gårding's (1970) results from a parallel EMG study lend support for a modelling of the word accents as positive, rather than negative pulses.[24] Furthermore, as mentioned above, by means of an interpretation of Meyer's (1937) data, Gårding (1970) argues that one of the positive pulses, i.e. the only one present in accent I, reflects sentence intonation.

An attempt to falsify Bruce's (1977) account of the word accents was made by Engstrand (1995, 1997).[25] The most obvious problem with Bruce's (1977) timing approach for Standard Swedish, as can be easily noticed informally, is that the $F_0$ fall it predicts for accent I is quite unstable in focal realisation, compared to the corresponding later pitch fall in accent II. Moreover, the fall in accent I can only occur if there is a pre-stress syllable.

In Engstrand's (1997) spontaneous speech data, accent II words were consistently produced with a (late) fall, as predicted both by the traditional approach and Bruce's (1977) timing model. In accent I words, when produced with an (early) fall as predicted by Bruce (1977), this fall had a substantially smaller range than the one in accent II words (mostly below 10 Hz); however, in the majority of the cases, there was in fact a *rise* instead of a *fall*. Engstrand's (1997) hypothesis is that the fall in accent I, if present, is an effect of sentence intonation, rather than a feature of the word accent. In a follow-up study with experimentally controlled speech, Engstrand (1995) elicited test sentences of the kind shown in (3), where capitals indicate a focal accent:

(3)      a.    Det var ju LÄNDERNA jag menade.
                  'It was the COUNTRIES/LOINS I meant.'
          b.    Det var dom STORA länderna jag menade.
                  'It was the LARGE countries/loins I meant.'

For both cases, two different versions were recorded, one with *länderna* as an accent I word (definite plural of *land* 'land'), the other with *länderna* as accent

---

[23]Öhman (1967), inspired by the Danish *stød*, in fact regards the word accents as variants of the glottal stop. Their assumed $F_0$-inhibiting function would be best supported if EMG data showed an *increased* activity of the *vocalis*, but a *decreased* activity of the *cricothyroid* muscle.

[24]Gårding (1970) found increased activity for both the *vocalis* and the *cricothyroid* muscle.

[25]But note that Bruce's (1977) account of the *sentence accent* is widely adopted, also by Engstrand.

II (definite plural *länd* 'loin'). In (3-a), *länderna* is produced with a focal accent and preceded by unaccented syllables, while in (3-b), *länderna* is unfocused and preceded by a focally accented accent II word, resulting in a high $F_0$ on the final syllable of *stora*, i.e. the syllable immediately preceding the initial stress in *länderna*. In summary, Engstrand (1995) found that (in a non-emphatic reading style), a fall in accent I was only present in the non-focal variant following the high pitch in *stora* (3-b). For accent II, a fall was found in both contexts, but the fall was significantly larger in sentence (3-b). Engstrand (1995) concludes that his previous hypothesis based on spontaneous speech is confirmed: The fall in accent I, if present, is a consequence of variations in sentence intonation (in this case, the directly preceding focal accent on *stora*), while in accent II, the fall is a consequence of the word accent itself. Engstrand (1995) also stresses the importance of perception for an adequate description of the data and suggests that, if there indeed is a positive specification for accent I, too, this could be determined by listening tests.

Bruce's (1977) study included two perception experiments in order to verify his analysis of the word accent contrast. The first one confirmed that the timing of a pitch fall is a sufficient cue to word accent in a phrase-final, post-focal position. The test phrase used was *INGA malmer*, where *malmer* is either an accent I word ('Inga Malmer', female name), or an accent II word ('no ores'), and the initial accent II word *INGA* (female first name, or plural of the negative, indefinite pronoun) is associated with a focal accent. This test context implies that there is a high pitch level in the syllable preceding the stressed syllable of *malmer* due to the focal accent on *INGA*, and a final fall due to the low phrase boundary. That is, by this experimental design it could be shown that a later timing of the pitch fall suffices as a cue to accent II. However, it could not show that accent I has a tonal specification, since the phrase intonation (i.e. the focal accent on *INGA* and the final boundary) would predict a fall even without the influence of any lexical tone specification.

By the same argument, however, Engstrand's (1995) test case (3-b) cannot falsify Bruce's (1977) observations concerning non-focal accents. For that, it would have been necessary to include even a sentence context which provides a non-focal word accent that is not immediately preceded by a focal accent, as for instance (4):

(4)         Det var de ANDRA stora länderna jag menade.
            'It was the OTHER large countries/loins I meant.'

However, the results of the second perception experiment conducted by Bruce (1977) cannot as easily be accounted for within a privative analysis. The test phrase used in the second experiment was *mellan målen* as opposed to *mellanmålen*, which may either be read as the two-word phrase 'between the meals' or

as the compound 'the between-meal snacks'. The two-word phrase consists of an accent II word (*mellan*) plus an accent I word (*målen*), while the compound is a single accent II word. The latter would usually contain two pitch peaks, one on *mel-* due to accent II, and one on *-mål-* due to the focal accent. The same peaks would be expected in the two-word phrase in case the second word (*målen*) were focally accented. However, in this case, Bruce's (1977) analysis would predict a third peak, medial in the phrase, located at the boundary between *mellan* and *målen* as a correlate of the final accent I word. The results showed that if such a third peak was present and sufficiently large, the test phrase was perceived as the two-word phrase. In absence of this medial peak, the phrase was perceived as the compound. Hence, an (early-timed) pitch peak seems to be a necessary cue even for an accent I word in order to be perceived as a prosodic word, at least in the case where it needs to be disambiguated from a compound. Engstrand's (1995) hypothesis predicts that a two-word phrase would not be tonally disambiguated from a compound, or, if it were, that the disambiguating tonal gesture would be a part of the phrase intonation. However, no proposals as to the nature of such a gesture are available.

The discussion so far has mainly been concerned with the phonetic manifestation of the word accents, and implications for phonological representation. However, the two perspectives that have been reviewed, the privative view and the equipollent view, share the assumption that there is an asymmetry between accent I and accent II, the former being the default category, the latter 'marked'. Other approaches, which are specifically concerned with the *lexical specification* of the word accents, have sometimes arrived at different conclusions. For instance, recently, Lahiri et al. (2005) have proposed that accent II is always post-lexical, while accent I is lexically specified in some cases.

In an ERP (Event Related Potentials) study, Roll (2009) investigated the processing of morphology and word accent in disyllabic inflected test words, consisting of monosyllabic stems plus suffix. Roll (2009) found an increased P600 effect, which indicates a processing difficulty, when a test word with a tonal pattern appropriate for accent I was combined with a suffix which induces accent II, like the indefinite plural *-ar* in the test word *minkar* 'minks'. However, no such increased P600 effect occurred in the opposite case, i.e. when a test word with a tonal pattern appropriate for accent II was combined with an accent I suffix like the definite singular suffix *-en* in *minken* 'the mink'. That is, a test word with an accent II suffix and an inappropriate tonal pattern is perceived as a mismatch, while a test word with an accent I suffix and an inappropriate tonal pattern is not, at least as far as we can tell from ERP data. This result lends strong support for the traditional privative account of the word accents (and against Lahiri et al. 2005), since 'accent II suffixes' seem to be specified for a specific tonal pattern, while 'accent I suffixes' are not.

In summary, it is clear that there is an asymmetry between accent I and accent II, which is evidenced by different sets of data, e.g. $F_0$ data on (focal) word accent realisation in spontaneous speech (Engstrand 1997) and ERP data on word accent and morphological processing (Roll 2009). In Riad's (2009) view, this should correspond to a markedness of the phonological representation, implying that the tonal pattern of accent I words (or accent II according to Lahiri et al. 2005) is determined by post-lexical factors. However, as discussed above, the results of Bruce (1977) are not easily explained within a purely privative approach. While it is true that accent I has an unstable tonal pattern in a *focal* context, the claim that *non-focal* accent I is similarly stable as non-focal accent II (Bruce 1977) has not been convincingly falsified yet. The lack of proposals on how the tonal pattern of non-focal accents should be accounted for in a strictly privative approach indicates a certain difficulty of doing so.

It seems, on the contrary, that it is easier to adopt the view that accent I is associated with an early pitch fall, and then to explain why accent I is much more unstable in a focal than in a non-focal context. It would seem that the accent I fall is reduced above all when its contribution to communication is redundant, which is in accordance with H&H-theory (Lindblom 1990). When a focal accent is present, the distinction between accent I and II is clearly provided by the timing of the focal accent (which is realised on the post-stress syllable in accent II), rendering the accent I fall less important. However, in special cases, e.g. when a two-word phrase is to be disambiguated from a compound, it might be more important to preserve the fall (cf. Bruce's 1977 second perception experiment). Also in non-focal position, where the additional cue from the timing of the focal accent is absent, it might be more important to preserve the falling pattern of accent I.

Thus, the available data so far support the 'split' perspective criticised by Riad (2009), but advocated, in some or the other way, by the account of Bruce (1977, 1998) as well as of Lahiri et al. (2005), namely to treat the issue of *phonological representation* (implying the issue of phonetic realisation) on the one hand, and the issue of *lexical specification* on the other hand as different phenomena.

### 3.3.2 Models of Swedish prosody

As in the case of German, a variety of models of Swedish prosody (accounting for intonation and word accents) have been proposed, e.g. Öhman (1967, 1968), Carlson and Granström (1973), Gårding and Lindblad (1973), Bruce and Gårding (1978), Fujisaki et al. (1993), Fant and Kruckenberg (2006). The present discussion of Swedish prosody is mainly based on the Lund model (Bruce 1977; Bruce and Gårding 1978; Gårding and Bruce 1981; Bruce and Granström 1993; Bruce et al. 1997; Bruce 2005; Bruce 2007), for the following reasons.

First, the development of the Lund model, in particular Bruce (1977), had

an important impact on phonological theory, more specifically the development of AM phonology (Pierrehumbert 1980; Goldsmith 1990) and is accepted as the state-of-the art of Swedish prosody in the international AM phonology community (cf. Ladd 1996; Ladd 2008; Gussenhoven 2004). Furthermore, unlike the modelling of the *word accents* (cf. 3.3.1), the assumptions made by the Lund model concerning Swedish *intonation* seem to be widely accepted also in the Scandinavian scientific community. For these reasons, the Lund model constitutes an appropriate point of departure. Second, following from the first point, many studies of Swedish intonation since Bruce (1977) have in one or the other respect been based on the Lund model.

### 3.3.3   Basic tonal patterns of Swedish in the Lund model

**Inventory and realisation of the basic tonal patterns**

The basic inventory of tonal categories of Swedish as assumed by the Lund model, formulated in terms of AM phonology, has already been displayed in Table 1.1 (Chapter 1). But some notes on the realisation and the phonotactics of these categories are in order.

Based on the analysis by Bruce (1977), the Lund model assumes that each prosodic word is associated with one of the two word accents. Furthermore, in each intonation phrase, normally one word is associated with a sentence accent (Bruce et al. 1998). Finally, each intonation phrase is closed by a final boundary tone.

As discussed in 3.3.1, both accent I and accent II are represented as a tonal fall from a H(igh) level to a L(ow) level. The distinction between the accents is a matter of the timing of this HL gesture: In accent II, the H is associated with the stressed syllable (H*+L)[26], while in accent I, the L is associated with the stressed syllable (H+L*), implying that the H is realised on the pre-stress syllable; the H is typically absent in utterances starting with an accent I stressed syllable.

The sentence accent (notated as H-)[27] is realised after the word accent gesture. In accent I, this results basically in a rise through the stressed vowel, which can be preceded by a fall (cf. 3.3.1): (H+)L* H-. In accent II, the resulting pattern is H*+L H-. That is, the sentence accent occurs after the stressed syllable. In simplex accent II words, the H- is usually located in the post-stress syllable, while

---

[26]An alternative notation is H*L. Here the H*+L notation is chosen since it is more comparable to the ToBI system, which is used for German.

[27]Note that the '-' indicates a *(sentence) accent* in the Swedish tradition, while in (G)ToBI, it marks an (intermediate) phrase boundary tone; in the simplified GToBI notation for German used in thesis, the '-' has been eliminated in order to avoid ambiguity. The '-' is used only to indicate a sentence accent in Swedish.

in compounds, the H- is associated with the secondary stress. In general, the exact location of the H- $F_0$ peak is highly context-dependent, and hence, the H- has been classified as a 'floating' tone (Bruce 1987).

Typically, the realisation of non-focal word accents (H+L* and H*+L) is different in the pre-focal part of a sentence than in the post-focal part. Pre-focally, each word accent is realised as an individual $F_0$ peak. Post-focally, the $F_0$ level of a preceding (focal or non-focal) accent is usually maintained until the next accent, which is then realised as a simple step down to a lower level, the timing of this step being the crucial distinction between accent I and II. That is, if the focal accent (H-) is realised late in a long sentence, several (pre-focal) $F_0$ peaks can precede the focal H-, while, if the focal accent occurs early in the sentence, the remainder of the sentence is realised as a series of (post-focal) $F_0$ *downsteps*.

A usual consequence of this post-focal type of interpolation between accents is a *hat pattern*, i.e. a high plateau connecting the focal accent and the first post-focal word accent. That is, this Swedish hat pattern is the default concatenation of a specific combination of accents (an H- plus an H*+L or H+L*), when these are not separated by a phrase boundary. Hence, the hat pattern can also be regarded as a *coherence* signal (cf. below). In this respect, the Swedish hat pattern is functionally similar to hat patterns in German (cf. 3.2.3).

Bruce (1977) originally investigated only statements with a falling intonation. The final boundary was hence modelled as an L tone (L%). Based on later studies on prosody in discourse, mainly based on spontaneous speech, a rise (LH%) as a second possible boundary tone has been introduced (e.g. Bruce et al. 2000). These and related studies have also lead to further enhancements of the model which concern global characteristics such as pitch range and register. The functions of such modifications are summarised below. First, however, another central feature of the Lund model is briefly summarised, namely its account of dialect typology.

### Dialect typology

Although this thesis is not concerned with dialectal variation as such, it is important to recognise that one of the main features of the Lund model is its account of a prosodic dialect typology for Swedish. This component of the model was already present in Bruce and Gårding (1978) and is further developed in e.g. Bruce (2005, 2007) and Bruce et al. (2007), in part based on extensive dialect data from the SWEDIA 2000 project (e.g. Bruce et al. 1999).

The crucial feature of the original dialect typology (Bruce and Gårding 1978) is the generalisation of the timing model for the word accents to all dialects of Swedish. That is, although word accent realisation and intonation differ largely between dialects, in every dialect, the word accent distinction is modelled as a matter of *relative* timing of an HL pattern. The distinction between the dialects

is, first, a matter of the *absolute* timing of the HL gestures. A second difference between dialects is the presence or absence of a separate tonal gesture for focus signalling. For instance, in South Swedish, focus is not signalled by an additional H- gesture as in Stockholm Swedish; moreover, the HL word accent patterns are timed later in South than in Stockholm Swedish; but in both dialects, the HL comes earlier for accent I than for accent II. It might be argued that this unifying approach of the word accent distinction itself is a further argument for the equipollent description of the word accents (cf. 3.3.1). Moreover, the typical argument against the equipollent perspective, i.e. that accent I is phonetically similar to sentence accents in German and English, may be true for Stockholm Swedish, but certainly not for South Swedish.

**Functions of the basic patterns and global modifications**

The Lund model of Swedish primarily accounts for *syntagmatic* aspects of intonation, namely the placement of sentence accents and boundary tones. Together, these intonational features are used to signal utterance-level prominence, exploited for focus signalling, and phrasing. Observe, however, that there is no simple mapping of intonational features on the one hand, and functions on the other hand. It has been reported that the placement of sentence accents also plays an important role for phrasing (e.g. Elert 1984; Bruce et al. 1993b).

In addition to the *local* aspect of placement of the basic intonational units (accents and boundary tones), the Lund model also accounts for variations of *global* parameters. In Gårding's early version of the model (Bruce and Gårding 1978; Gårding 1979; Gårding 1984b),[28] global sentence intonation was modelled as a *tonal grid*, which serves as the reference for accentual targets. Different sentence intonations, e.g. for the signalling of information structure (theme–rheme) or speech act or sentence type distinctions (statement–question), are modelled as variations of the tonal grid. The grid can be compressed or extended, falling (typical for statements) or rising, or divided into several parts, e.g. a rising and falling one.

Several studies on the relation between discourse or dialogue structure and prosody in Swedish have been undertaken with the purpose of further developing the Lund model, in part also in a speech technology context (e.g. Bruce 1982b; Bruce et al. 1990; Bruce and Touati 1992; Bruce et al. 1993a; Bruce et al. 1993b;

---

[28] According to Gårding (1987) there have actually been two different models of Swedish intonation in Lund, or two different versions of the Lund model. Both are based on Bruce (1977) in the sense that they apply the idea of tonal targets for accents and boundaries. The crucial aspect of the first one, advocated by Eva Gårding, presented in Bruce and Gårding (1978) and further developed in e.g. Gårding (1979, 1984a), is the tonal grid, which has not been adopted in Bruce's further developments of the Lund model (e.g. Bruce 1982a).

Bruce et al. 1996; Bruce et al. 1997; Bruce et al. 1998; Bruce et al. 2000). The general results of these studies can be summarised as follows: First, basically the same patterns are used in spontaneous speech as the ones observed earlier in laboratory speech (Bruce et al. 1990). Second, global features such as overall pitch range, pitch register, or baseline slope, are varied for the signalling of coherence and boundaries between intonation phrases (e.g. Bruce 1982b; Bruce et al. 1990; Bruce et al. 1993b; Bruce et al. 1998).[29] Another typical correlate of coherence, or the absence of a boundary, is the *hat pattern* (cf. above). Coherence and boundary signalling, in turn, are related to discourse- or dialogue structure. For example, several successive phrases can be produced with a continuous downdrift of the $F_0$ baseline, and without resetting range and register, in order to signal topical coherence between the phrases.

Other findings in connection with phrasing in spontaneous speech were the recognition of different levels of phrasing (minor and major boundaries; Bruce et al. 1998), as well as different *initial* (%H vs. %L) or final junctures (LH% vs. L%; Bruce et al. 2000). These additional boundary patterns constitute an exception to the otherwise purely *syntagmatic* nature of the basic pattern inventory of the Lund model, since they imply a *paradigmatic* choice between two patterns (L% vs. LH% in the case of the final junctures). However, these additional patterns have not been investigated systematically from a functional perspective yet; it has only been stated that the final rise (LH%) can be used to signal continuation (Gussenhoven 2004).

In summary, the Lund model assumes that Swedish intonation patterns are built up of a rather small set of distinctive units – a focal accent and set of boundary tones. In addition, global features are varied for several discourse-related purposes.

### 3.3.4 Some further studies on Swedish intonation

In the previous section, the basic tonal patterns of Swedish as assumed by the Lund model have been presented. The review of studies presented there was restricted to studies directly concerned with the development of the Lund model. The goal of the present section is to complement the review of Swedish intonation research with a brief overview of some further studies. Some of these have also been based on the Lund model, but not directly concerned with its further development.

The first large-scale instrumental study of intonation in connected speech in a

---

[29]That is, the idea of a tonal grid (cf. Gårding 1984b), although usually without explicit reference to this term, has been re-introduced into the later versions of the Lund model (e.g. Bruce and Touati 1992).

variety of Swedish was undertaken by Hadding-Koch (1961). The study was pioneering, among others because of its methodology, using a relative large corpus, combining auditory analyses with acoustic measurements, and complementing with controlled perception experiments using synthetic speech. Since Hadding-Koch's (1961) study is concerned with South Swedish, which differs prosodically from the variety investigated in this thesis to a great extent, it is not treated in detail here. It can be noted, however, that the study does not recognise any patterns that are not accounted for by the current Lund model, with the exception of a three-fold contrast of phrase final patterns: *rise*, *fall*, and *sustain*.

A study explicitly and exclusively concerned with intonation, while keeping word accent constant, is reported in Bredvad-Jensen (1981). In her material, a test sentence consisting of three accent II words, is elicited in a controlled manner with focal accent on either of the three words, like in Bruce (1977). However, two further sentence- or utterance-related variables are included, namely sentence mode (declarative, interrogative) and a three-fold contrast of attitude (neutral, polite, and determined). Again, since the investigation is concerned with South Swedish, the results will not be discussed in detail here. To give a short summary, Bredvad-Jensen (1981) found that several tonal parameters interact in the expression of the different functional dimensions, such as the $F_0$ range in connection with the sentence accent (which is expanded both for the signalling of questions and politeness)[30], the pre-focal and post-focal contour (which is e.g. compressed in questions), the overall $F_0$ level (e.g. higher in questions), and the phrase-final contour (a rise being found in questions, a fall in statements). The interplay of attitude and sentence mode is somewhat complex. For instance, while politeness is expressed by an expanded range for the sentence accent for both sentence modes, determination gives rise to compression only in questions, not in statements.

Heldner (2001) concentrates on one specific unit of the basic inventory of the Lund model, namely the focal accent, and studies additional, or alternative, acoustic correlates beyond $F_0$. As a general result, focally accented words are not only marked by the typical $F_0$ rise (H-), but also by increased duration and spectral emphasis. Moreover, Heldner (2001) concludes that the $F_0$ rise alone is neither necessary nor sufficient to signal that a word is in focus, and hence, that although it is the most important correlate, $F_0$ must interact with other cues. Some of Heldner's (2001) studies are discussed in more detail in later sections.

The studies reviewed in this section so far have provided insights which are important for a comprehensive understanding and modelling of Swedish intonation, not least since further functional dimensions, such as attitudes, have been taken into account. However, the investigations did not have crucial implications

---

[30]Bredvad-Jensen (1981) relates *question* and *politeness* to the 'weak' meaning category, and *statement* and *determination* to the 'strong' category in the sense of Cruttenden (1979).

for the inventory of basic patterns to be assumed for Swedish. A recurring result is that basic patterns can be modified (e.g. in terms of $F_0$ range) for the expression of various functional distinctions.

However, more recent studies have given rise to the assumption that further basic units of Swedish intonation need to be assumed beyond those present in the Lund model. A first example is a series of studies by House (e.g. 2002, 2003, 2005) on the signalling of interrogative mode. The traditional account for Standard Swedish (Gårding 1979) has been to describe interrogative intonation as a raising of the topline (in the tonal grids of Gårding 1979) and a widening of the focal accent $F_0$ range. A final rise in questions has typically been regarded as optional in Swedish, and its $F_0$ excursion has been characterised as moderate compared to the rise found in e.g. English (Hadding-Koch and Studdert-Kennedy 1964; Gårding 1989).

In a perception experiment, House (2002) found that also a *delay* of the focal accent peak makes an important contribution to perceived question intonation. In particular, each of the two parameters *widened range* and *peak delay* could give rise to a slight majority of question votes (about 60%), but together they resulted in a clear impression of question intonation. House (2002) states that a late peak position results in a rise in the initial part of the final vowel (the final word of the test phrase used was the accent II word *flyga* 'to fly', where the focal accent is realised on the final syllable), and that this rise in itself is probably a perceptual cue to interrogativity. Hence, "[t]he fact that the timing of the final focal accent in Swedish can be used to signal question intonation adds an extra dimension to the function of the focal accent" (House 2002). It should be stressed that, in the experiment reported in House (2002), even for the most delayed peak position in the stimulus series, a short final fall is preserved. House (2002) does not comment on the perceptual relevance of this short fall, which has a reported duration of 25 ms for the latest peak position (stimulus 6). It is plausible to assume that the primary perceptual effect of stimulus 6 is a rise throughout the entire final vowel, perhaps slightly dampered during the last 25 ms; it is at least unlikely that a final fall is perceived in a similar manner as for the earlier peak positions.[31]

In further analyses of spontaneous speech and additional perception experiments, House (2005) found further evidence for the usage of a delayed focal accent peak in questions. Moreover, he relates the distinction of a late focal peak (resulting in a final rise) and an early[32] peak (which is followed by a final fall)

---

[31]The range of the fall is not reported exactly, but can be estimated from the graphic presentation of the stimuli in House (2002) to be approximately 6.2 or 9.8 semitones (st) for the series with narrow or wide range respectively. The resulting rates of $F_0$ change for the duration of 25 ms are 248 st/s or 392 st/s, respectively, i.e. much larger than the observed values for the *fastest* rates found in natural speech (between 59 and 96 st/s for falling $F_0$ according to Xu and Sun 2002).

[32]This "early peak" is not to be confused with the German *early peak* (Kohler 1991a), or the

in a question to a difference in dialogue act. A late peak (i.e. a final rise) is hypothesised to signal that the intention of a question is primarily *social interaction*, while an early peak (i.e. a final fall) is rather used in genuine *information* questions. House (2005) found support for this hypothesis in a series of perception experiments, and related the phenomenon to the *frequency code* (cf. 2.4). A result of House (2005) is thus also a revision of the status of the final rise in questions – it is obviously not optional, but instead implies a crucial meaning difference. Unlike in House (2002), in the perception studies in House (2005), the final focal accent was realised on a final lexically stressed syllable (of the monosyllabic, and hence accent I, word *du* 'you').

House (2002, 2003, 2005) treats the final rise as the result of a delayed focal accent peak. However, the rising pattern described by House (2005), as found in the spontaneous speech data, is strikingly similar to the pattern that would be expected in a corresponding communicative context in German (Kohler 2005b). Thus, an alternative interpretation of House's (2005) results would be that we are dealing with a rising boundary pattern and, in fact, with a *low sentence accent*, which can contrast paradigmatically with the well-established *high* one (H-). That is, even though not stated by House (2002, 2003, 2005) himself, his results suggest that we might need to assume a further basic unit of Swedish intonation.

Finally, two explicit proposals concerning previously unrecognised units of Swedish intonation have been made recently, both concerning initial boundary phenomena. It has been observed earlier (Bruce 1982a) that an utterance containing at least two prosodic words (or stress groups) often contains two sentence accents (referred to as *phrase accents* in Bruce 1982a), one associated with the initial word, the other with the focussed word. Bruce (1982a) explained this phenomenon by assuming that these utterances consist of two prosodic phrases, each of which is associated with a phrase accent; this phrase accent *can* serve the additional purpose of signalling *focus*. A crucial feature of the initial H- commonly observed is that its peak is reached *later* than in an H- that signals focus. This late-timed initial sentence accent has been further investigated from a dialogue perspective by Horne (1991), Hansson (2000), and Horne et al. (2001). Hansson (2000, 2001) suggested that the late-timed initial H- might be associated with the signalling of *contrastive topics* (cf. 2.3.2), as discussed in more detail in 6.1.2.

Recently, Myrberg (2009) has proposed that the initial sentence accent and the focal accent proper should be conceptually distinguished, although they share the same tonal representation. She suggests the term *initiality accent* (IA) and argues that this IA differs in several respects – function, distribution, and realisation –

---

basic pattern type *early fall* discussed in 3.2.3 (H+L*). The "early peak" in House's (2005) terms is characterised by a relative early occurrence of the H- peak; it is thus phonetically more similar to the *medial peak* discussed for German.

from a *focal accent* (FA). The main function of the IA is to signal the beginning of a new phrase. That is, Myrberg's (2009) proposal is to recognise a further unit related to the *syntagmatic* aspect of Swedish prosody.

All proposals concerning the initial rising $F_0$ pattern reviewed so far have treated this late-timed rise as a sentence *accent*, more specifically as a realisational variant of the H- (Bruce 1982a; Horne 1991; Hansson 2000; Horne et al. 2001; Myrberg 2009). An alternative interpretation as an initial *boundary tone* has recently been proposed by Roll (2006) and Roll et al. (2009). Based on a production study (Roll 2006) and a perception study using ERP (Roll et al. 2009), Roll suggests that the function of this boundary tone is primarily syntactic, marking the left edge of main clauses. We will briefly return to the discussion of initial phenomena in connection with the results presented in Chapter 6.

To summarise, numerous studies have been conducted on Swedish intonation, some of which explicitly in connection with the Lund model development. However, with few exceptions, no further basic phonological units have been proposed for the modelling of Swedish intonation beyond the inventory represented in the Lund model. The exceptions concern the *syntagmatic* level (the IA proposed by Myrberg 2009 and the initial boundary tone proposed by Roll 2006). However, it has been argued that the results of House (2005) indicate a possible need to assume a *paradigmatic* contrast of utterance-level accents.

## 3.4 Swedish vs. German intonation

The previous two sections have provided a brief overview of German intonation (3.2) and a review of some studies on Swedish intonation and word accents (3.3), both with the main focus on the inventory of basic tonal patterns. These sections have indicated that typical descriptions of German and Swedish intonation, based on different traditions, have used different conceptualisations of intonation. Two key notions which illustrate this difference are the *nucleus* on the one hand, and *focus* on the other hand. Their relation is discussed in Section 3.4.1, and based on that, Section 3.4.2 compares the *basic nuclear patterns* of German and Swedish as they can be derived from the reviews in 3.2 and 3.3.

### 3.4.1 Focus and the nucleus

While *focus* has played a role in both works on Swedish (e.g. Bruce 1977) and German intonation (e.g. Altmann et al. 1989; Uhmann 1991; Féry 1993), the *nucleus* concept has also been central in many approaches to German and other West Germanic languages (e.g. O'Connor and Arnold 1973; Gussenhoven 1984; Féry 1993), but not in the Swedish tradition. Instead, the most comparable notion used

in the Swedish tradition is the *focal accent*, which is tightly connected to functional notion of *focus*. What is the difference, and what are the similarities between the concepts of *focus* and *nucleus*? Why have different conceptualisations of intonation been preferred in German and Swedish? And what is the relation between these notions and the observed difference in the assumed intonational systems of Swedish and German? The following sections discuss some possible answers to these questions.

**Form vs. function**

A crucial difference between *nucleus* and *focus* is that the former notion is anchored in prosodic form, and the latter in prosodic function, since the *nucleus* refers to the 'most prominent' accent in an intonation phrase (cf. 2.1.2), and *focus* is a notion of *information structure*, approximately referring to the 'non-presupposed information' (but cf. Section 2.3.1). The fact that one notion has played a more salient role in the German, and the other in the Swedish tradition may have several reasons. One reason is probably the presence or absence of a word accent distinction in Swedish or German, respectively, which provides a language-specific difference in point of departure, as discussed in the following paragraphs.

   Although also focus and other functional concepts have played a role in many works on German intonation, it can be stated that, historically, intonational *form* has been the point of departure in the German tradition, since early works often had a pedagogical purpose (e.g. von Essen 1964). Of course, the *function* of intonation was not ignored by these works – foreigners learning German should not only be able to produce the correct patterns, but also to use them in the right context. However, German intonation has traditionally been treated from a form-first perspective (cf. 2.5), implying that a variety of intonational forms is observed, and then, it is attempted to grasp the functions, or the meaning distinctions, associated with these patterns. As a result, many functional descriptions of German intonation patterns can be regarded as post-hoc characterisations, where several different semantic and pragmatic dimensions are merged, as illustrated in Table 3.1 above. The function-first perspective, which was also applied to German (e.g. Altmann et al. 1989), but to a lesser extent, would select a certain semantic or pragmatic parameter at a time, and investigate how this parameter is manifested in intonation (e.g. focus and sentence mode in the case of Altmann et al. 1989).

   While thus for the German tradition prosodic *form* has been the typical point of departure, the Swedish tradition may be characterised by a *functional* point of departure, provided by the word accent distinction. Hence, one of the primary research questions concerning Swedish prosody has always been and still is the question of how the word accents are manifested phonetically, reflecting a proto-

typical function-first approach. The quest for the tonal manifestation of the *word* accents has also led to the recognition of a *sentence* accent (Bruce 1977), and hence to the secondary research goal of disentangling the influence of word accent and sentence accent on the surface melodic pattern. It can be noted that the term *sentence accent* does not carry the same functional connotations as the later term *focal accent*. Indeed, *sentence accent* is also a common term in the German tradition. Moreover, in the first works on Swedish where a *sentence accent* was mentioned (e.g. Meyer 1937) or explicitly modelled (e.g. Öhman 1967), the account of the sentence accent was primarily formal, since no comprehensive understanding of its usage had emerged yet. This changed, however, with the work of Bruce (1977), who related the sentence accent to the concept of *focus*.

To summarise, there seems to be a traditional orientation towards intonational *form* as a point of departure in the German (i.e. West Germanic) tradition, but towards *function* in the Swedish tradition. It would seem that a function-first perspective is more likely to develop in the context of research on languages with lexical tonal phenomena, since they have a function that is easily accessible for native speakers – the potential of distinguishing between words – in contrast to most intonational functions (cf. Xu and Xu 2005 for similar discussion in connection with English and Mandarin). This difference in research traditions – *form-first* vs. *function-first* – is even, as outlined above, reflected in the conceptual distinction *nucleus* vs. *focus*.

In 2.2.1, we have considered the possibility that *syntagmatic* intonational contrasts (related to accent and boundary placement) have above all *intrinsic* functions, exploited for the basic prosodic functions of prominence signalling and grouping, while *paradigmatic* contrasts (such as choice of accent type) often have *relational* functions. Moreover, we have discussed that relational functions are typically quite elusive. It would thus not be surprising if relational functions were generally taken into account to a lesser extent in classical function-first approaches, as in the Swedish tradition, compared to form-first approaches. This might be one explanation for the fact that so far no *paradigmatic* distinctions of sentence accents have been recognised for Swedish. The studies presented in this thesis will, as mentioned in 2.5, follow a function-first approach, but also attempt to include some relational functions, as outlined in 3.5.

**Nuclear accent vs. focal accent**

Nevertheless, there are similarities in the concepts *nucleus* on the one hand, and *focal accent* on the other hand. Both the nuclear accent and the focal accent typically represent the most prominent accent of an intonation phrase. Furthermore, the *nuclear accent* is, of course, also tightly connected to the signalling of *focus*, since in German, as in Swedish, the focussed constituent is associated with the

most prominent accent of the phrase (e.g. Gussenhoven 1984; Féry 1993).

The different *pragmatic usages* of (semantic) focus in terms of Krifka (2007), as discussed in 2.3.1, may provide a further link between focus and the nuclear accent. The discussion and modelling of nuclear accents (e.g. in the context of German intonation research) typically involves both the *syntagmatic* and the *paradigmatic* dimension of intonation, i.e. both *accent placement* and the *choice of accent type*. As suggested in 2.3.1, different *pragmatic usages* of focus (Krifka 2007) might be related to the choice of nuclear accent type.

### Pre-nuclear accents vs. non-focal accents

For both the *focal accent* and the *nuclear accent*, the characterisation as the most prominent accent in an intonation phrase implies that in case the phrase contains further accents, these must be less prominent. For Swedish, two tonal prominence levels are generally distinguished: Besides the *focal accent*, which always co-occurs with a *word accent* gesture[33] there are *non-focal* accents, whose tonal patterns are determined by the word accent only. As discussed in 2.1.2, the domain of a word accent is assumed to be the *prosodic word* (Bruce 1998). That is, a word accent highlights a certain foot (the one containing the stressed syllable) within a (prosodic) word, but it does not, according to the model, highlight the word within the phrase. Moreover, in Standard Swedish, the distinction between the two tonal prominence levels (focal vs. non-focal) is rather clear-cut phonetically, because focal accents are marked by an additional tonal gesture (H-). Hence, the assumption of two tonal prominence levels seems to follow naturally from the structure of Swedish.

In the case of German, a similar distinction between *nuclear* and the less prominent *pre-nuclear* accents is often made (e.g. Féry 1993; Grice et al. 2005; Baumann 2006). This is motivated, because obviously not all accents potentially contained in a German phrase signal an individual focus. However, there is, of course, a crucial difference between pre-nuclear accents of German and non-focal accents of Swedish, since in German, which lacks a tonal function connected to the word, all accents (i.e. pre-nuclear and nuclear) must in some way be regarded as related to the utterance level. Furthermore, the phonetic differences between pre-nuclear and nuclear accents in German is not as clear as between Swedish non-focal and focal accents. It is thus not surprising that not all approaches of German intonation agree upon whether pre-nuclear accents should be regarded as different from nuclear accents concerning their form and prominence level. For instance, in the Kiel Intonation Model (Kohler 1991a), no formal difference between pre-nuclear and nuclear accents is assumed, apart from the fact that the final accent

---

[33]At least, this is assumed by the Lund model, the alternative view being that only accent II contributes a word-related tonal pattern, cf. 3.3.1

of a phrase is combined with a phrase-final pattern (corresponding to boundary tones in AM phonology). In GToBI, the nuclear accent is treated as the most prominent accent, but again, there is no formal difference between pre-nuclear and nuclear accents: All accents proposed by GToBI are allowed (when labelling a data corpus) in any position in a phrase. In contrast, Féry (1993) assumes a clear formal difference between nuclear and pre-nuclear patterns, reflected in a reduced inventory of pre-nuclear accents and a potential for tonal reduction.[34]

A further difference between non-nuclear accents in German and non-focal accents in Swedish is that non-focal accents may occur before (pre-focally) or after (post-focally) the focal accent. In other words, the focal accent does not imply that words occuring after the focussed word are deaccented. In contrast, a nuclear accent of German is typically treated as the most prominent *and the last* accent of a phrase, and hence, only pre-nuclear, but no post-nuclear accents are assumed. That is, using a nuclear accent for signalling focus implies a post-focal deaccentuation in German, but not in Swedish.

## 3.4.2 Basic nuclear patterns of German and Swedish

A preliminary comparison of the tonal inventory of German and Swedish as assumed by contemporary accounts (GToBI and the Lund model) was presented in Table 1.1 in Chapter 1. The table simply listed all tonal patterns associated with accents and final boundary tones assumed by the two models. Table 3.2 provides a revised comparison, restricted to basic *nuclear* patterns. That is, the patterns displayed in Table 3.2 describe the tonal contour from the nuclear accent of the phrase to the final boundary. For German, the patterns included in Table 3.2 are not based on GToBI alone (as is the case for Table 1.1), but on a general overview of German intonation models (Section 3.2); in fact, the inventory of five basic patterns suggested in Section 3.2 and Table 3.2 is almost identical to Féry's (1993) proposal. Concerning Swedish, Table 3.2 is based on the assumption that a focally accented word is comparable to a word carrying the nuclear accent in German, as discussed in 3.4.1.

Table 3.2 indicates several differences between German and Swedish. First, the tonal pattern of a nuclear accent in Swedish is structurally more complex, since

---

[34] According to Féry (1993), only two of her five nuclear accents can also occur in pre-nuclear position, namely H*L and L*H, which can be linked with the nuclear pattern, resulting in mono-tonal surface realisations, H* or L*, of pre-nuclear accents. Féry's (1993) perspective of a reduced inventory of pre-nuclear as compared to nuclear accents is shared by many studies within AM phonology on German as well as other European languages (e.g. Atterer and Ladd (2004) and references therein), which are concerned with the tonal alignment of "the" pre-nuclear rise (L*H in Féry's (1993) terms), implying that there is only one. Niebuhr and Ambrazaitis (2006), based on spontaneous speech data, argue for a distinction of two types of rises (approximately L+H* and L*+H in GToBI) even in pre-nuclear position.

Table 3.2: A comparison of basic nuclear intonation patterns, derived from Section 3.2.3 for German (represented in simplified GToBI) and from 3.3.3 for Swedish (Lund model notation).

| German | Swedish |
|---|---|
| H+L* L% | |
| H* L% | w.a. H- (w.a.) L% |
| L*+H L% | |
| H* LH% | w.a. H- (w.a.) LH% |
| L* LH% | |
| word accent (w.a.) = H+L* │ H*+L | |

the focal accent co-occurs with a word accent. Second, the nuclear accent may be followed by post-nuclear accents, which then consist of the word accent pattern only (cf. 3.3.3 for the realisation of post-focal accents in Swedish). Third, if we disregard the tonal complexity introduced by the *word accents*, a rather simple system of *intonation* patterns has been assumed for Swedish, as compared to the system of German (or other West Germanic languages). In particular, there is only one sentence accent (H-), while for German, a paradigm of accents has been assumed, four of which are included in Table 3.2 (H*, L*, H+L*, L*+H). When combined with final boundary tones, the German inventory contains five basic nuclear patterns, while only two different nuclear patterns can be derived from the Lund model for Swedish.

This observation is equivalent to Gussenhoven's (2004) statement quoted in Chapter 1. However, the difference in the intonational repertoires displayed in Table 3.2 may have several reasons. It might, of course, reflect a *factual* difference in the intonation systems of German and Swedish. In this case, the presence of the word accent might *directly* explain the relatively simple intonation system of Swedish, but as discussed in Chapter 1, this explanation does not seem to be fully satisfactory, since there are tonal languages with more complex intonation systems than the one assumed for Swedish (cf. Yip 2002; Gussenhoven 2004).

Hence, it might also be the case that the difference in intonation systems indicated in Table 3.2 is a result of the observed difference in research tradition (cf. 3.4.1), and thus only *indirectly* related to the presence of the word accents in Swedish. If the difference in pattern inventories of Swedish and German assumed so far really were, at least to some degree, a result of different research traditions, this would imply that the difference between Swedish and West Germanic intonation is not necessarily as large as reflected by the contemporary models. In fact,

the results of some studies reviewed in 3.3.4, the most relevant for the present discussion being House (2005), indicate that further tonal categories might be required for a more complete description of Swedish intonation. To conclude, in order to be able to compare the intonational systems of Swedish and German, it would seem necessary to apply the same research perspective to the two languages in a systematic manner.

## 3.5  Research questions and overview of studies

The general research question of this thesis has been outlined in Chapter 1, but can be refined based on the background presented in Chapters 2 and 3 as follows:

- Does Standard Swedish have a similar inventory of basic *nuclear* intonation patterns as German (or the West Germanic languages in general), including a *paradigmatic* contrast of utterance-level accents?

The question can also be asked from a functional perspective:

- How does Swedish express those intonational *meanings* that are expressed by different nuclear patterns in German?

Since Swedish has a word accent distinction, a further question is:

- How are the word accents manifested in connection with other utterance-level accents – if such can be established – than the traditional focal accent (H-)?

Both production and perception data are reported in this thesis. The perception study (7.3) is designed in order to test a specific hypothesis derived from the third production study (7.2), and is described in more detail in Chapter 7. The production studies partly build upon each other, but the general questions asked in the production studies are derived from the background given in this chapter, as discussed in the following paragraphs. The general method adopted for the production studies is a function-first approach (cf. 2.5).

The general idea underlying the first production study (Chapter 5) is to investigate how the intonational functions or meanings that have been associated with the five basic nuclear patterns of German (cf. 3.2) are expressed in Swedish. However, the review on German has shown that the form–function link in German intonation is rather complex, or at least difficult to determine, and hence, the hypothesised form–function relations (cf. 3.2.3) cannot be taken for granted. Therefore, even German data need to be included in the investigation and consequently, the investigation receives a comparative component, too.

In this first study (Chapter 5), an attempt is made to elicit the five basic nuclear patterns of German by means of adequate material which is constructed in advance and then acted out by a number of speakers. Nine different test contexts (conditions) are used. Corresponding materials in Swedish are created and the patterns produced by Swedish speakers are compared with those produced by the German speakers. The material for Swedish in this first study contains only accent I.

Based on the results of the first study (Chapter 5), a more concrete version of the general hypothesis (cf. Chapter 1) can be formulated, namely that Swedish, besides the high, or (late) rising utterance-level accent ((H+)L* H- for accent I), also has a low, or (early) falling one (H+L* L- for accent I). This accent occurred in several conditions, among them a *confirmation*, where also in German an early falling pattern (H+L*) was expected and found.

The second production study (Chapter 6) provides a first test of this hypothesis. The study concentrates on two of the conditions used in Chapter 5 (*confirmatitve response* and *new-information response*) and adds a further pragmatic dimension, namely *narrow focus* location. Hence, two dimensions of focus signalling are included in the study: the choice of *accent type* related to the *pragmatic usage* of focus (a *'plain'* vs. a *confirmative* usage of focus, cf. 2.3.1) and *accent placement* related to focus location (*narrow focus* on either of three constituents). The idea is that, if the falling pattern found in a confirmation (H+L* L-) includes a *sentence accent* (L-), then it should differ formally and functionally from a similar falling pattern (H+L*) which, does not represent a sentence accent, but rather a (non-nuclear) *word accent*. The specific hypothesis is that the rising-falling nuclear pattern ((H+)L* H- L%) of Swedish is used to signal a *'plain' focus*, rendering *new* information, while the falling pattern (H+L- L%) signals a *confirmative focus*, related to *given* information.

A secondary research question of Chapter 6 is the signalling of a further information structure category, namely a *contrastive topic* (cf. 2.3.2), as discussed in more detail in 6.1.2.

The third production study (Chapter 7) is concerned with Swedish only and investigates how the hypothesised L- accent interacts with the word accent. In this study, again, confirmations and new-information responses are elicited, but unlike in the second study, focus location is kept constant. Instead, word accent is varied systematically. A major finding of this study is that the word accent distinction may be optionally *neutralised* in connection with the low L- accent produced in confirmations. The processing of word accent neutralisation in a confirmation is further investigated in a final perception study, using reaction time measurements (7.3).

# Chapter 4

# Database and methods

## 4.1 Overview

The thesis contains four empirical studies, three production studies analysing acoustic data, and one perception study based on reaction time measurements and listener judgements. The purpose of this chapter is two-fold. First, sections 4.1.1 and 4.1.2 provide an overview of the empirical basis of the entire thesis, excluding, however, methodological details which concern the individual studies only; such details are presented together with the results of the corresponding study in Chapters 5 to 7.

Second, this chapter serves as a reference for the three production studies presented in Chapters 5, 6, and 7. The speech corpora used in the production studies are presented in detail in this chapter (4.2, 4.3, and 4.4), since first, some properties are shared by the corpora, and second, some details are not directly relevant for the presentation of the results in later chapters. The present chapter also describes the $F_0$ normalisation method used in Chapters 5 and 6.

### 4.1.1 Perception data

The perception study (section 7.3) is a follow-up study of the production study presented in 7.2 and comprises a reaction time (RT) experiment. This experiment was run under experimental conditions in different locations (KTH Stockholm; Stockholm University; Lund University), and results from 20 subjects were collected. Details on the method are presented in 7.3.

### 4.1.2 Production data

Three different corpora were designed and recorded, which provide the basis for the three production studies (Chapters 5, 6, and 7), cf. the outline in 3.5:

- Corpus A: Exploring nuclear patterns in German and Swedish (Ch. 5)

- Corpus B: Focus, topic, and accent type in German and Swedish (Ch. 6)

- Corpus C: Nuclear pattern type and word accents in Swedish (Ch. 7)

While corpora A and B were recorded with a Swedish and a German group of nine speakers each, corpus C contains nine Swedish speakers only. For each language, corpora A and B were recorded by the same groups of speakers and in the same recording session. Corpus C was recorded independently by another group of (Swedish) speakers, although this group overlaps to a large extent with the Swedish speaker group for A and B, as displayed in Table 4.2 below.

For all corpora, subjects were asked to read, or act out, test sentences or phrases as parts of constructed dialogue situations. The instruction was always to render the text as natural and appropriate to the context as possible, although without being too theatrical. The precise arrangement and presentation of the test material, i.e. the elicitation and recording procedure, differed in the different corpora, cf. 4.2 and 4.3.

In particular, three different strategies were applied for the elicitation, in short: (i) 'reading from a context description' (used for a subset of corpus A), (ii) 'simulated dialogues with pre-recorded questions' (B and a subset of A), and (iii) 'simulated dialogues read by pairs of speakers' (C). The remainder of this chapter presents the three corpora A, B, and C in more detail.

## 4.2   Corpus A and B

### 4.2.1   Materials and test conditions

Corpora A and B were designed in order to elicit test phrases in a variety of contexts which differ in some pragmatic parameters and to compare the melodic patterns produced in these contexts by German and Swedish speakers. That is, the corpora contain a German and a Swedish version of the test material. The main criterion for the choice of linguistic material for Corpora A and B was thus to ensure that the German and Swedish versions were as comparable as possible from a segmental and prosodic perspective. Therefore, the material is based on Swedish–German cognates with similar phonetic realisations and identical lexical stress patterns. Furthermore, all Swedish words in this material have accent I. As presented in more detail below for each corpus separately, the entire material of Corpora A and B is built up of only three content words, namely *Wallander* (a sur-

name), *November* ('November'), and (German/ Swedish:) *verlängert/ förlänger*[1] ('continue', 3rd person singular, present tense).

A second criterion was to avoid voiceless consonants within words in order to provide uninterrupted $F_0$ tracks. At the same time, the material should provide robust acoustic landmarks at certain segment boundaries in order to support a reliable spectrographic segmentation of the material.

A third criterion was to ensure that the test phrases or sentences would be semantically plausible. Since the task of the speakers was to simulate natural conversation, this criterion should probably be regarded as relevant for providing a high degree of *validity* of the elicited speech data, while the second criterion is related to the *reliability* of the study. Although it might be the case that speakers could manage to express pragmatic distinctions by means of prosody in a natural way even with semantically odd material, the use of such material might imply unnecessary risks. However, when the first two conditions are to be met, it is not entirely trivial to satisfy even the third criterion. Hence, the test sentence chosen for Corpus B (introduced below) might seem odd at a first sight, which is related to the fact that the test sentence lacks a direct object, although the verb *förlänger/ verlängert* is canonically transitive. However, in a colloquial style and given a context where the (missing) direct object represents *given* information, the construction is entirely natural in both German and Swedish. Therefore, the naturalness of the material was ensured by designing appropriate situational frame contexts for the test phrases. These contexts also served to simulate the pragmatic conditions to be investigated, as explained in more detail in the following sections.

**Corpus A**

Corpus A contains two test phrases, a primary and a secondary one, in a Swedish (S) and a German (G) version each, given orthographically and in a broad phonetic transcription in (1):

(1)  S  1. *i november* [i nʊˈvɛmbəɹ]; 2. *Wallander* [vaˈlandəɹ]
     G  1. *im November* [ɪm noˈvɛmbɐ]; 2. *Wallander* [vaˈlandɐ]

        1. 'in November', 2. 'Wallander' (surname[2])

The corpus was designed in order to elicit the five basic nuclear patterns of German discussed in 3.2.3, and to test which patterns would occur in Swedish in corresponding contexts. For some of the patterns, two or three possible contexts

---

[1]This cognate pair contains a difference in consonant quantity between the German and the Swedish version (cf. 3.1), which is, however, not relevant for the conclusions drawn from the study where it is used (Chapter 6).

[2]police inspector in Henning Mankell's novels; famous in Sweden and Germany

were included in the corpus (discussed in more detail in 5.1), resulting in nine test contexts, or conditions, in total. For each condition, a situational context is provided, but two different types of conditions have been designed, referred to as type (i), comprising six conditions, and type (ii), comprising three conditions.

The contexts designed for the six conditions of type (i) are listed in (2) to (7). In each condition of type (i), there is a context description followed by the short text (printed in bold face) which was to be read aloud by the subject. This text consisted of the test phrase ('in November' in five cases, 'Wallander' in one case), embedded is an utterance frame (e.g. 'Fine. In November.').

(2)     SUMMARY: concluding summary

    S     *Du diskuterar tillsammans med dina korridorskompisar när er fest ska äga rum. Efter en stund beslutar ni att ha festen i november. Du avrundar ämnet genom att säga:* **Bra. I november.**

    G     *Du diskutierst mit Deinen Mitbewohnern, wann Eure Party stattfinden soll. Nach einer Weile entscheidet ihr Euch für den November. Du schließt das Thema mit folgenden Worten ab:* **Gut. Im November.**

       'You are discussing a possible time for your party with your residence mates. After a while, you decide to have the party in November. You close the discussion by saying: Fine. In November.'

(3)     EXCLAM: exclamation

    S     *Det uppstår ett missförstånd: Alla pratar om november, bara du tror att det handlar om september. Slutligen fattar du:* **Jaså! I november! Nu förstår jag.**

    G     *Es entsteht ein Missverständnis: Alle reden vom November, nur Du denkst, es geht um den September. Schließlich begreifst Du es:* **Ach so! Im November! Jetzt versteh ich das.**

       'There is some misunderstanding: Everyone is talking about November, but you believe that it's about September. Finally, you realise: Ok! In November! Now I understand.'

(4)     SURPRIS: surprised feedback

    S     *Din studiekamrat Johan berättar för dig att han lämnade in sitt examensarbete i november. Du är mycket imponerad, eftersom sista datum för inlämning var först i januari och ingen brukar bli klar tidigare. Du är förbluffad:* **Wow! I November! Inte dåligt!**

    G     *Dein Studienkollege Johan erzählt Dir, dass er im November seine Examensarbeit abgegeben hat. Du bist sehr beeindruckt, denn der Abgabetermin war erst im Januar, und normalerweise schafft es keiner,*

*früher fertig zu werden. Du staunst:* **Wow! Im November! Nicht schlecht!**

'Your fellow student Johan tells you that he has submitted his thesis in November. You are very impressed, because the deadline is in January, and normally nobody manages to get finished earlier. You are astonished: Wow! In November! Not bad!'

(5)  QREPET: repetition question

S  *Din granne berättar för dig att han snart ska ha en viktig tenta. Du tror att han nämnde november, men du är inte säker på det. Du frågar om:* **Och när skulle din tenta äga rum? I november?**

G  *Dein Nachbar erzählt Dir, dass er bald eine wichtige Prüfung hat. Du glaubst, dass er den November erwähnt hat, aber Du bist Dir nicht sicher. Du vergewisserst Dich:* **Und wann ist die Prüfung? Im November?**

'Your neighbour is telling you that he'll have an important exam soon. You believe that he has mentioned November, but you are not sure. You ask again: And when does the exam take place? In November?'

(6)  QDISBEL: disbelieving question

S  *Din kompis Martin berättar för dig att han ska ha en viktig tenta i november. Du är förvånad, eftersom din syster läser samma ämne som Martin, och hon ska ha tenta först i december. Du undrar alltså:* **Är du säker på det? I november?**

G  *Dein Freund Martin erzählt Dir, dass er im November eine wichtige Prüfung hat. Du bist verwirrt, denn Deine Schwester studiert dasselbe Fach wie Martin, und ihre Prüfung ist erst im Dezember. Du fragst also:* **Bist Du sicher? Im November?**

'Your friend Martin is telling you that he will have an important exam in November. You are puzzled, because your sister is studying the same subject as Martin, and her exam is scheduled in December. So you ask: Are you sure? In November?'

(7)  ADDREQ: addressing for request

S  *Du vill be din kollega Wallander om hjälp:* **Wallander? Skulle jag kunna få be dig om en tjänst?**

G  *Du willst Deinen Kollegen Wallander um Hilfe bitten:* **Wallander? Dürfte ich Sie um einen Gefallen bitten?**

'You would like to ask your colleague Wallander for help: Wallander? May I ask you for a favour?'

In the conditions of type (ii), there is a context description like in type (i), which is, however, shorter than for type (i) and referred to as *frame context*. In addition, there is a *context question* prior to the test phrase. The *frame context* is the same for all three conditions of type (ii) and is displayed in (8).

(8)     Frame context for NEWINFO (9), CORRECT (10), and CONFIRM (11)

> S     *Du är polis och träffar en gammal kollega.  Ni småpratar lite om jobbet.*
>
> G     *Du bist Polizist und triffst einen alten Kollegen. Ihr redet ein bisschen über die Arbeit.*
>
> 'You are a police officer meeting a former colleague. You are talking about the job.'

The context questions and the corresponding answers (i.e. the test phrases embedded in an utterance frame) for the three conditions of type (ii) are listed in (9) to (11).

(9)     NEWINFO: new-information response

> S     *När ska du egentligen ta semester, då? – **I november.***
>
> G     *Wann machst Du denn eigentlich Urlaub? – **Im November.***
>
> 'When are you going on holiday, actually? – In November.'

(10)    CORRECT: corrective response

> S     *Och du tar din semester alltså i oktober igen, då? – **Nej! I november.***
>
> G     *Und Du nimmst Deinen Urlaub dann also wieder im Oktober? – **Nein! Im November.***
>
> 'And you're going on holiday in Oktober again, right? – No! In November.'

(11)    CONFIRM: confirmative response

> S     *När skulle du egentligen ta semester?  I november, eller?  – **Ja, i november.***
>
> G     *Wann machst Du eigentlich nochmal Urlaub? Im November, oder? – **Ja, im November.***

> 'When are you going on holiday, actually? In November, right? – Yes, in November.'

These context presentations also show that the primary test phrase 'in November' was used in eight of the nine conditions, while the secondary phrase 'Wallander' was used only in the ADDREQ context. The reason for the introduction of the secondary phrase was simply that the ADDREQ context requires a name or some other reference to a person in order to be meaningful.

## Corpus B

For corpus B, one test sentence was constructed in a Swedish (S) and a German (G) version, given orthographically and in a broad phonetic transcription in (12):

(12)  S  *Wallander förlänger till november.*
         [ʋaˈlandə fœˈlɛŋːə tʰɪl nʊˈʋɛmbəɪ]
      G  *Wallander verlängert bis November.*
         [ʋaˈlandɐ fɛˈlɛŋɐt bɪs noˈʋɛmbɐ]

      'Wallander is continuing until November'

The test sentence was elicited in seven conditions by means of constructed dialogue contexts. The structure of the contexts was the same as type (ii) in Corpus A. Each dialogue context consisted of a situational frame context, which was the same for all conditions, and a unique context question. The frame context is given in (13) and the context questions are listed in (14) to (20).

(13)    Frame context for Corpus B
      S  *Du är polis och träffar en gammal kollega. Ni pratar om pensioneringen och om möjligheten att förlänga sin tjänst.*
      G  *Du bist Polizist und triffst einen alten Kollegen. Ihr redet über die Pensionierung und die Möglichkeit, seinen Dienst zu verlängern.*

         'You are a police officer meeting a former colleague. You are talking about retirement and the possibility to continue working.'

One condition was designed in order to investigate the signalling of an information structure context referred to as *contrastive topic plus focus* (CTOP+FOC), cf. 2.3.2. For this condition, the test sentence was embedded into an utterance frame, as shown in (14). In (14) to (20), the test sentence (12) is represented by three dots (...).

(14)    CTOP+FOC: contrastive topic plus final focus

    S    *Och vet du redan fram till när dina kollegor förlänger?*
        *– **Alltså, . . . Det är allt jag vet.***

    G    *Und weiß Du schon bis wann Deine Kollegen verlängern?*
        *– **Also, . . . Das ist alles, was ich weiß.***

    'And do know already how long your colleagues will continue work-
    ing? – Well, ... . That's all I know.'

The remaining six conditions are defined as combinations of two pragmatic di-
mensions: *narrow focus location* and the *referential status* of the focussed mate-
rial. That is, the test sentence was elicited with *narrow focus* on the first, second,
or third content word, in each case both as *new (plain focus)* and *given* information
*(confirmative focus)*. In the case of a confirmation, the test sentence was preceded
by *ja* ('yes').

(15)    NEWINFO-1: initial plain focus (new information)

    S    *Och vem förlänger till november? – . . .*
    G    *Und wer verlängert bis November? – . . .*

    'And who is continuing until November? – . . . '

(16)    CONFIRM-1: initial confirmative focus (given information)

    S    *Men vem var det som förlänger till november, då? Det var väl Wal-
        lander, eller? – **Ja, . . .***

    G    *Aber wer verlängert denn bis November? Das war doch Wallander,
        oder? – **Ja, . . .***

    'But who is continuing until November? It's Wallander, isn't it? –
    Yes, . . . '

(17)    NEWINFO-2: medial plain focus (new information)

    S    *Vad är det då som Wallander gör fram till November? – . . .*
    G    *Was macht Wallander denn bis November? – . . .*

    'What's Wallander doing until November? – . . . '

(18)    CONFIRM-2: medial confirmative focus (given information)

    S    *Men vad var det som Wallander gör fram till November? Han
        förlänger väl, eller? – **Ja, . . .***

    G    *Aber was macht Wallander bis November? Er verlängert doch,
        oder? – **Ja, . . .***

'But what's Wallander doing until November? He's continuing, isn't he? – Yes, – ...'

(19)  NEWINFO-3: final plain focus (new information)

S  *Och fram till när förlänger Wallander? – ...*
G  *Und bis wann verlängert Wallander? – ...*

'And until when is Wallander continuing? – ...'

(20)  CONFIRM-3: final confirmative focus (given information)

S  *Fram till när förlänger Wallander egentligen? Till november, eller?*
   *– Ja, ...*
G  *Bis wann verlängert Wallander eigentlich? Bis November, oder?*
   *– Ja, ...*

'Until when is Wallander continuing, actually? Until November, right? – Yes, – ...'

## 4.2.2  Elicitation and recording procedure

German and Swedish speakers were asked to read the test material from a computer screen in response to the provided context. For Corpus A, two different strategies for the presentation of the context were used in two different subsets of the material.

For subset (i) of Corpus A, in each trial, the entire context description for a given condition, including the test phrase in its utterance frame (cf. (2) to (7)) was simply displayed on the screen. The subject's task was to read the provided context silently and to try to imagine the situation described in the context, and when ready, to render the test text (printed in bold face (2) to (7)) in a normal conversational style.

For subset (ii) of Corpus A, as well as for Corpus B, in the beginning of each trial, only the frame context ((8) or (13)) was displayed on the computer screen in written form. After the speaker had clicked a button, a pre-recorded context question was played to her/him via headphones, and simultaneously, the test phrase appeared on the screen. Even here, the task was to try to imagine the situation described in the frame context, and then, based on the written context and the additional auditive context question, to render the test phrase in a normal conversational style. In all corpora and subsets, the subjects were allowed to repeat each trial until they were satisfied.[3]

---

[3]Generally, this self-monitoring procedure worked successfully; it was hardly ever necessary

Prior to the recordings, it was explicitly pointed out for the subjects that many trials would be similar or even identical. They were encouraged to try to regard each trial as it if had occurred for the fist time. All instructions were given both orally and in written form.

The material of Corpus A and B was randomised, and each test item occurred five times in the recording session. The list of test items also included further conditions not reported on in this thesis. The recording session was divided into two blocks. The first block comprised all cases where strategy (ii) was used (i.e. Corpus B and a part of Corpus A), while the second block comprised strategy (i) (i.e. the remainder of Corpus A). This division was undertaken since, first, the two strategies required different instructions (e.g. only in (ii) the speakers were required to use headphones), and second, it enabled the speakers to take a break from the recordings.

Most speakers completed the first block in approximately 20–25 minutes and the second block in 10–20 minutes. The relatively larger variation in duration of the second block (strategy (i)) is probably related to the fact that the task in this block contained more written context information and no auditive context information at all, which some speakers found more demanding; in addition, the speakers probably differed in reading tempo. A complete recording session for one speaker, including instructions and a pause between the two blocks, usually took 60–90 minutes.

All but three speakers were recorded in an experimental studio at the Humanities Laboratory, Lund University, using a *Shure BG 4.0* microphone. The data collection was performed using the BAS SpeechRecorder (Draxler and Jänsch, URL). This computer program organised the recording session and presented the contexts and test sentences to the subjects. For that, an external computer screen, a mouse, headphones, and the microphone were placed in the recording studio, but the recording device (a mixer and a laptop computer) and the investigator were located in a separate control room.[4] The recordings were digitised at 44.1 kHz and 24 bit and saved directly onto hard disk.

The context questions that were presented to the speakers auditively in the elicitation of Corpus B and subset (ii) of corpus A were pre-recorded by a 28-year-old male speaker from Kiel, and a 30-year-old male speaker from Linköping. The two speakers represent the Standard varieties of German and Swedish. The context questions were recorded in the anechoic chamber at the Humanities Laboratory, Lund University, using a *Neumann U87 Ai P48* microphone, digitised at

---

to interrupt the subjects.

[4]Three German subjects were recorded in the anechoic chamber at the Institute of Phonetics and Speech Processing at the University of Munich, using a *Neumann TLM 103* microphone. Note that also they were speakers of Standard German. For these recordings, the recording equipment (mixer and laptop computer) could not be placed in a separate room, for technical reasons.

44.1 kHz and 24 bit and saved directly on hard disk.

## 4.2.3 $F_0$ normalisation

The $F_0$ data obtained from corpora A and B were time and register normalised. The purpose of this normalisation was to support visual comparison and presentation of the data, and to make it possible to calculate mean $F_0$ contours across several repetitions of the same intonation patterns produced by different speakers. This is useful in order to illustrate the general characteristics of the dominant patterns found for the test conditions.

The method applied for the normalisation is similar to the method developed within the FK (Fant and Kruckenberg) approach (e.g. Fant et al. 2002; Fant and Kruckenberg 2004, 2006) and the method used by Xu and colleagues (e.g. Xu 1999; Liu and Xu 2005), however, with a modification inspired by Modulation Theory (MDT; Traunmüller 1994; 2005). Both the Xu and the FK method involve a log scaling (e.g. in semitones) of $F_0$ values, as well as a time normalisation. The time normalisation is achieved by representing chunks of speech (e.g. syllables) by a certain number $n$ of temporally distributed $F_0$ measurements, and hence, depending on the choice of $n$, it involves a more or less detailed stylisation of the original $F_0$ contour. The method developed here and its relation to the approaches by Xu and FK is explained in more detail in the remainder of this section.

First, the utterances were manually segmented, using spectrograms and wave form diagrams in Praat (Boersma and Weenink URL), into segments of varying size (e.g. representing phones, syllables, or words), depending on the purpose of the study (cf. Chapters 5 and 6). Second, the automatic pulse marking provided by Praat, which was used for the calculation of $F_0$, was manually corrected. Errors in the automatic pulse marking are common, especially when the voice quality deviates from modal, but also in the case of rapid changes in $F_0$. In most of these cases, periods are still easily detectable in the wave diagram by visual inspection, and hence, missing or erroneously placed pulses can be manually corrected. This procedure prevents missing $F_0$ values as well as typical $F_0$ errors such as octave jumps. Of course, $F_0$ values will still be misleading in extreme cases of creaky voice. These are common in utterance-final position, and in the present corpora, more frequent for Swedish than for German speakers. Moreover, $F_0$ was automatically trimmed by an algorithm that removes spikes and sharp edges from the $F_0$ courses (Xu 1999).

Third, $F_0$ was time normalised by extracting $n$ temporally equidistant $F_0$ measurements from each segment. Different resolutions were used in the different studies in this thesis, namely 10 measurements per phoneme-sized unit in Chapter 5, but only 10 measurements per word (and 10 measurements per syllable for a subset of the data) in Chapter 6. Even in the case of the least detailed res-

olution, the results are near-natural (i.e. smooth) but time-normalised copies of the original contours. This degree of detail is similar to the one applied in the studies by Xu and colleagues (e.g. Liu and Xu 2005), and higher than in the FK approach (e.g. Fant and Kruckenberg 2006), where only one or two measurements per syllable are used. These first three steps were performed using the Praat script _TimeNormalizeF0 (Xu URL).

The corrected, trimmed, and time-normalised $F_0$ data were originally extracted using a Hz scale. Thus finally, these data were speaker normalised by converting the measurements into semitones, and at the same time rescaling them so that a speaker's base $F_0$ value $F_b$ would roughly correspond to 0 on the semitone scale. For that, an estimation of $F_b$ was used as the reference value $F_r$ in the calculation of semitones for each speaker. Semitones are calculated according to equation 4.1, where $\Delta F$ represents the transformed $F_0$ value in semitones, $F_r$ the reference value in Hz, and $F_m$ the original $F_0$ measurement in Hz to be transformed.

$$\Delta F = 12 \times log_2 \frac{F_m}{F_r} \qquad\qquad (4.1)$$

Thus, as in the FK approach, a semitone scale is used. It has been reported in various studies that a semitone scale more adequately represents pitch patterns in speech from a perceptual perspective than other available psycho-acoustic scales such as mels, Bark, and ERB-rate (Traunmüller and Eriksson 1995b; Nolan 2003).

In the FK normalisation, an arbitrary, speaker-independent reference value $F_r$ of 100 Hz is used, and depending on the purpose of the normalisation, the data are rescaled so that 0 on the semitone scale corresponds to the speaker-specific mean $F_0$.

In the present studies, an arbitrary reference value is inappropriate, since mean contours are plotted across different groups of speakers (cf. Chapters 5 and 6). When using an arbitrary and fixed value for $F_r$, register differences between two mean curves would not necessarily depend on the simulated pragmatic condition, but potentially also on the voice register of the individual speakers included in the two samples. Therefore, a speaker-specific reference was required, and the speakers' base value $F_b$ was preferred over the mean, which was motivated by MDT (Traunmüller 1994, 2005).

The $F_0$ base value $F_b$ can be defined as the $F_0$ baseline on which the $F_0$ minima of an utterance are located (Traunmüller 2005). That is, due to declination, $F_b$ is slightly higher at the onset of an utterance than at the end (Traunmüller 2005), but approximately, $F_b$ is equivalent to the level reached at the end of an utterance with falling intonation (Traunmüller and Eriksson 1995a). In MDT, $F_b$ ($F_{0c}$ in Traunmüller 2005) is the $F_0$ of the carrier signal, i.e. the unmodulated voice. The carrier signal depends on extralinguistic factors (age, sex, vocal effort, etc.), and hence, $F_b$ can vary even within a speaker (e.g. due to changes in vocal effort).

However, $F_b$ is hardly affected by conventional, both linguistic and paralinguistic, $F_0$ modulation (Traunmüller and Eriksson 1995a). For instance, a higher degree of liveliness, e.g. due to the type of discourse, usually results in an increased $F_0$ standard deviation ($sd$), which in turn is achieved mainly by raising local $F_0$ maxima, while the minima, and in particular the utterance-final $F_0$ in cases of falling intonation, are almost kept at a constant value (e.g. Bruce 1982a; Menn and Boyce 1982; Liberman and Pierrehumbert 1984). As a consequence, the mean $F_0$ increases with increased $sd$ caused by variation in liveliness, as confirmed by the data of Johns-Lewis (1986) and Graddol (1986), cf. also Traunmüller and Eriksson (1995a) for an overview of these data. Hence, $F_b$, approximately representing the $F_0$ minima, is more robust and a better representation of a speaker's voice (or the carrier signal) than the mean. Lindh and Eriksson (2007) provide further experimental evidence for the robustness of a baseline measure as compared to the mean.

Traunmüller and Eriksson (1995a) suggest a formula for the calculation of $F_b$ which is based on the mean and the standard deviation of $F_0$. However, the mean $F_0$ represents a speaker's $F_0$ distribution only if a normal distribution can be assumed, but typically, there is some positive skewness in an $F_0$ distribution (Jassem et al. 1973). Lindh (2006) uses the median instead of the mean for the calculation of $F_b$, since the median is less effected by outliers (octave jumps).

For the present study, the $F_b$ was estimated neither based on the mean nor the median, since the present data provide a possibility to estimate $F_b$ more directly, using only five utterances per speaker taken from Corpus B, namely the five repetitions of the test sentence as uttered with initial plain focus (NEWINFO-1). In this condition, all speakers, both Germans and Swedes, have produced approximately the same melodic pattern: a rise-fall on the first word, followed by a low plateau or a fall on the second word, and low $F_0$ on the last content word of the utterance. Despite extensive voice quality effects on the last word for some speakers, it was qualitatively judged that the measurements oscillated around a plausible $F_b$ of that speaker. Therefore, the average $F_0$ of the final word was calculated for each token, and the mean of these averages across all five repetitions was used as an estimation of the speakers' $F_b$. These estimated $F_b$ values for the speakers involved in corpora A and B are included in Table 4.2 below.

It must be noted that this procedure for estimating $F_b$ is highly dependent on the structure of the particular speech material of the present study, and thus not generally applicable. However, according to Traunmüller and Eriksson's (1995a) calculations based on published $F_0$ statistics from a variety of studies, the average $F_b$ for speakers of European languages are 93.4 Hz (male) and 163.8 Hz (female). The estimates made for the 18 speakers of the present study show considerable inter-speaker as well as inter-language variation (cf. Table 4.2), but the average values for male (97.2 Hz) and female (173.8 Hz) speakers for both

languages pooled are, although slightly higher, similar to the averages reported by Traunmüller and Eriksson (1995a). One-sample t-tests fail to show a significant difference between our values and the values from Traunmüller and Eriksson (1995a): $t(7) = 0.8, p > .05$ (male); $t(9) = 1.62, p > .05$ (female).

## 4.3   Corpus C

Corpus C differs in several properties from Corpora A and B. First, since the corpus was designed in order to investigate the Swedish word accent distinction, there is only a Swedish version of this corpus. Second, as in Corpus A and B, context questions were used in order to elicit the conditions. However, for Corpus C, the test dialogues were not presented to a speaker by a computer program, but instead they were read by pairs of subjects.

### 4.3.1   Materials and test conditions

All test dialogues had the general structure displayed in (21), where A and B represent the two speakers.

(21)    A    <context-question(s)>
        B    *ja, det/den är/var* <target-phrase>
             [ɑː de/dɛn e/ʋa] <target-phrase>

             'yes, it is/was <target-phrase>'

The <target-phrase> consisted of a disyllabic target word with lexical stress on the first syllable, preceded by a monosyllabic function word. Ten different target words were used, five with accent I and five with accent II, which are displayed in Table 4.1. The accent I class contained five nouns in definite singular form, while the accent II class contained the same words in indefinite plural form.

A phonetic and a semantic criterion were applied simultaneously for the composition of the corpus. First, microprosodic effects were largely controlled by choosing words (i) with close (stressed) vowels only, and (ii) so that perturbations due to the initial consonant should be counter-balanced in the corpus (cf. the consonants preceding the stressed vowel). Second, the chosen words are rather common, i.e. they can be expected to occur frequently in every-day conversation (*bov* being perhaps an exception).

The content of <context-question(s)> is listed in (22) to (31). It was designed individually for each target word, in order to provide a situational context that was as natural as possible, and at the same time, to elicit the test sentence (bold face

Table 4.1: The 10 target phrases used for Corpus C.

| Accent I | Accent II |
|---|---|
| *med bilen* [me ˈbiːlən] 'by car' | *med bilar* [me ˈbiːlaɹ] 'with cars' |
| *om boven* [ɔm ˈbuːʋən] 'about the villain' | *om bovar* [ɔm ˈbuːʋaɹ] 'about villains' |
| *på stigen* [pʰo ˈstiːɡən] 'on the trail' | *om stigar* [ɔm ˈstiːɡaɹ] 'about trails' |
| *på stolen* [pʰo ˈstuːlən] 'on the chair' | *om stolar* [ɔm ˈstuːlaɹ] 'about chairs' |
| *med kniven* [me ˈkniːʋən] 'with the knife' | *till knivar* [tʰɪl ˈkniːʋaɹ] 'for knives' |

in (22) to (31)) as a new-information response (NEWINFO), or as a confirmative response (CONFIRM).

(22)   ***Ja, det var med bilen.*** 'Yes, (it was) by car.'

NEWINFO   *Vet du egentligen hur vi skulle åka till Helsingborg i morgon?* 'Do you actually know how we were going to get to Helsingborg tomorrow?'

CONFIRM   *Hur skulle vi egentligen åka till Helsingborg i morgon? Det var väl med bilen, eller?* 'How were we actually going to get to Helsingborg tomorrow? By car, right?'

(23)   ***Ja, det var med bilar.*** 'Yes, (it was) with cars.'

NEWINFO   *Vad tyckte Lasse egentligen mest om att leka med? Minns du det?* 'What did Lasse actually like best to play with? Do you remember that?'

CONFIRM   *Vad var det han tyckte mest om att leka med. Det var väl med bilar, eller?* 'What was it he liked best to play with? With cars, right?'

(24)   ***Ja, den var om boven.*** 'Yes, it was about the villain.'

NEWINFO   *Skrev inte Bosse en D-uppsats om en person i den senaste Wallanderromanen? Minns du vem den handlade om?* 'Didn't Bosse write a term paper about a character in the latest Wallander novel? Do you remember who it was about?'

CONFIRM   *Handlade inte din D-uppsats om en person i den senaste Wallanderromanen? Den var väl om boven, eller?* 'Didn't your term paper deal with a character in the latest Wallander novel? It was about the villain, right?'

(25)   ***Ja, den var om bovar.*** 'Yes, it was about villains.'

NEWINFO   *Du berättade nån gång att Åke hade skrivit en D-uppsats*

*som hade med kriminalromaner att göra. Minns du vad den hand-*
*lade om?* 'You told me once that Åke had written a term paper
which was dealing with crime stories.  Do you remember what is
was about?'

CONFIRM    *Var det inte din D-uppsats som hade med kriminalromaner*
*att göra? Den var väl om bovar, eller?* 'Wasn't it your term paper
that was dealing with crime stories? It was about villains, right?'

(26)    ***Ja, det var på stigen.*** 'Yes, (it was) on the trail.'

NEWINFO     *Hur skulle vi egentligen ta oss till andra sidan skogen?*
*Minns du det?* 'How were we going to get to the other side of the
forest, actually? Do you remember that?'

CONFIRM     *Hur skulle vi egentligen ta oss till andra sidan skogen? Det*
*var väl på stigen, eller?* 'How were we going to get to the other side
of the forest, actually? On the trail, right?'

(27)    ***Ja, den var om stigar.*** 'Yes, it was about trails.'

NEWINFO     *Du sa nån gång att Sven hade en bok om vandring eller*
*så. Minns du exakt vad den handlade om?* 'You mentioned once
that Sven had a book about hiking or something. Do you remember
exactly what it was about?'

CONFIRM     *Du sa nån gång att du har en bok om vandring eller så.*
*Den var väl om stigar, eller?* 'You mentioned once that have a book
about hiking or something. It was about trails, wasn't it?'

(28)    ***Ja, det var på stolen.*** 'Yes, (it was) on the chair.'

NEWINFO     *Minns du var du har lagt kokboken nånstans?* 'Do you re-
member where you put the cook-book?'

CONFIRM     *Du sa att du hade lagt kokboken nånstans. Det var väl på*
*stolen, eller?* 'You said that you put the cook-book somewhere.  It
was on the chair, right?'

(29)    ***Ja, den var om stolar.*** 'Yes, it was about chairs.'

NEWINFO     *Du berättade nån gång att Åke hade skrivit en D-uppsats i*
*möbeldesign. Minns du vad den handlade om?* 'You told me once
that Åke had written a term paper in furniture design.  Do you re-
member what is was about?'

CONFIRM     *Du berättade nån gång vad din D-uppsats i möbeldesign*
*handlade om. Den var väl om stolar, eller?* 'You told me once what
your term paper in furniture design was about.  It was about chairs,
wasn't?'

(30)    ***Ja, det var med kniven.*** 'Yes, (it was) with the knife.'

> NEWINFO   *Vet du hur vi skulle skära steken?* 'Do you know how we should slice the roast?'
>
> CONFIRM   *Hur sa du jag skulle skära steken? Det var väl med kniven, eller?* 'What did you say about how I should slice the roast? With the knife, right?'

(31)     **Ja, den är till knivar.** 'Yes, it's for knives.'

> NEWINFO   *(i köket) Vet du vad de använder den här lådan till?* '(in the kitchen) Do you know what they use that box for?'
>
> CONFIRM   *(i köket) Vad använder ni den här lådan till? Den är väl till knivar, eller?* '(in the kitchen) What you do use that box for? It's for knives, right?'

## 4.3.2   Elicitation and recording procedure

For corpus C, the speakers received the dialogues printed on paper and were recorded in pairs. They were instructed to read the dialogues in a normal conversational style. Moreover, they were encouraged to discuss their readings and, if necessary, to repeat any dialogue until they were satisfied.[5] This method is adopted from Kohler and Niebuhr (2007) and has the advantage that a near-spontaneous speaking style is approached, while a high degree of experimental control is maintained.

Prior to the recordings, it was explicitly pointed out for the subjects that many trials would be similar. They were encouraged to try to regard each trial as it if had occurred for the fist time. All instructions were given orally.

The 20 test dialogues (10 target words, 2 situations) were randomized and mixed with 43 further dialogues (not reported on here), yielding a corpus of 63 dialogues in total. The dialogues were arranged so that each speaker would read the A-part in every second dialogue. The speakers read the whole corpus twice, with interchanged parts on the second run, so that effectively, each speaker read the whole corpus once. One recording session consisting of instructions, the two runs, and a break in between took approximately 1h 15min.

The subjects were recorded in pairs, each one sitting in an experimental studio at the Humanities Laboratory, Lund University, communicating via the recording microphones *(Shure BG 4.0)* and headphones. The recording equipment (mixer and computer) and the investigator were located in a third, separate room. The recordings were digitised at 44.1 kHz and 24 bit and saved directly onto hard disk. Since 10 speakers were recorded in total (cf. 4.4), there were five recording sessions.

---

[5]Generally, this self-monitoring procedure worked successfully; it was hardly ever necessary to interrupt the subjects.

## 4.4   Speakers

Table 4.2 provides an overview of the 21 speakers reported on in this thesis. Five additional speakers (one for Swedish, corpus C; four for German, corpus A and B) were recorded but excluded from the analyses for different technical reasons.

Corpus A and B contain recordings by nine speakers of each language (German and Swedish). Corpus C contains data from nine Swedish speakers. As the table shows, there is a considerable overlap between the two sets of Swedish speakers. The recording sessions for Corpus A/B took place 14 months (speakers SF3, SF4, SF5, SM2, SM3) or 19.5 months (speaker SM1) after the sessions for Corpus C.

Among both language groups, the speakers have slightly different regional backgrounds, but all can be classified as belonging to the varieties under investigation, i.e. Standard Northern German, and Standard (Sveamål) Swedish.

Table 4.2: The 21 speakers recorded for the production studies. For speakers of Corpora A and B, an estimate of $F_b$ is included, cf. 4.2.3. Corpora A and B were recorded 14–20 months after Corpus C.

| Speaker id | Language | Sex | Age | Est. $F_b$ (Hz) | Corpus |
|:---:|:---:|:---:|:---:|:---:|:---:|
| GF1 | German | female | 24 | 180.5 | A, B |
| GF2 | German | female | 26 | 209.9 | A, B |
| GF3 | German | female | 35 | 187.0 | A, B |
| GF4 | German | female | 28 | 167.1 | A, B |
| GF5 | German | female | 45 | 186.8 | A, B |
| GM1 | German | male | 26 | 85.7 | A, B |
| GM2 | German | male | 36 | 114.3 | A, B |
| GM3 | German | male | 30 | 77.4 | A, B |
| GM4 | German | male | 24 | 109.3 | A, B |
| SF1 | Swedish | female | 41 | 181.0 | A, B |
| SF2 | Swedish | female | 27 | 174.0 | A, B |
| SF3 | Swedish | female | 30; 31 | 151.8 | C; A, B |
| SF4 | Swedish | female | 31; 32 | 149.4 | C; A, B |
| SF5 | Swedish | female | 25; 26 | 150.3 | C; A, B |
| SF6 | Swedish | female | 38 | – | C |
| SF7 | Swedish | female | 29 | – | C |
| SM1 | Swedish | male | 22; 24 | 85.8 | C: A, B |
| SM2 | Swedish | male | 50; 51 | 94.9 | C; A, B |
| SM3 | Swedish | male | 38; 39 | 100.7 | C; A, B |
| SM4 | Swedish | male | 31 | 109.3 | A, B |
| SM5 | Swedish | male | 25 | – | C |

# Chapter 5

# Exploring nuclear patterns in German and Swedish

## 5.1 Introduction

It has been argued in Chapter 3 that the *syntagmatic* aspects of intonation have played a dominant role in studies and models of Swedish intonation so far, while in the case of German, also the *paradigmatic* dimension of intonation has traditionally been recognised to a great extent. The syntagmatic aspects are related to the signalling of prominence relations and phrasing, two basic functions of intonation which play an important role for the structuring of utterances and discourse, including aspects of the marking of information structure (cf. 2.3). However, the intonation patterns of two utterances can be identical from a syntagmatic perspective, while they still differ in meaning due to a paradigmatic intonational contrast. For example, two utterances consisting of the words *a b c d* may be produced with identical phrasing and prominence patterns, e.g. *(A) (b c D)*, where word *A* and *D* are accented, and there is a phrase boundary between *A* and *b*. Given this identical syntagmatic structure in the two utterances, the melodic pattern associated with the accents and phrase boundaries may still be different – say, in utterance 1, *D* is produced with a low accent and a phrase-final rise, while in utterance 2, *D* is produced with a high accent and a final fall.

### 5.1.1 Goal

This first study is concerned only with paradigmatic intonational contrasts, or more specifically, with the paradigm of *nuclear* patterns. Nuclear patterns consist of the last, and usually most prominent, accent of an intonation phrase plus the phrase-final intonation, cf. 3.2.2. As discussed in 3.2, it is convenient to assume

Table 5.1: A comparison of basic nuclear intonation patterns, derived from 3.2 for German (represented in simplified GToBI) and from 3.3 for Swedish (Lund model notation). Modified version of Table 3.2.

|  | German | Swedish |
|---|---|---|
| *early fall* | H+L* L% | |
| *medial fall* | H* L% | H- L% |
| *late fall* | L*+H L% | |
| *fall-rise* | H* LH% | H- LH% |
| *simple rise* | L* LH% | |

an inventory of (at least) five different nuclear pattern types for German, listed in Table 5.1. In the case of Swedish, however, the Lund model (e.g. Bruce and Gårding 1978; Bruce and Granström 1993) accounts for two nuclear pattern types only: a fall following the focal accent (H- L%)[1], or a *fall-rise* (H- LH%). Table 5.1 also contains descriptive labels for the pattern types which are inspired by the literature on German intonation (cf. 3.2).

The goal of this thesis is to test whether the Swedish inventory of intonation patterns is more similar to the inventory of German than what is indicated by the Lund model (cf. 3.3.3). This chapter presents an initial, exploratory step of the investigation. For that, the study attempts to elicit the five basic nuclear patterns of German in a group of German speakers by means of speech material, which is designed on the basis of the existing functional hypotheses for these patterns (cf. 3.2.3), and applies a Swedish version of the same material to a group of Swedish speakers. In order to exclude the possibility of syntagmatic intonational variation in this study, i.e. variations in accent placement or phrasing, the material of this study is limited to utterances containing a single content word.

## 5.1.2   Eliciting the basic nuclear patterns of German

It is evident from the overview in 3.2.3 that the functional hypotheses on the five German patterns span a wide spectrum of pragmatic dimensions, including, for instance, the signalling of speech acts or sentence mode, information structure as

---

[1]The classification of the Swedish H- L% in Table 5.1 has been guided by the label, whose tonal composition is the same as the German H* L%. It could be argued that the pattern is phonetically more similar to a *late fall* in German, when the contribution of the word accent (accent I in this study) is taken into account: (H+)L* H- L%. The results of this study show that the Swedish pattern is phonetically similar to both the *medial* and the *late fall* of German.

well as the speaker's attitude towards the hearer and the message. The overview in
3.2.3 also shows that a variety of functions have been suggested for each pattern,
which can even be seemingly contradictory, illustrating the high degree of context-
dependency of the functions of the patterns. It is obviously difficult to pinpoint
the functions of each pattern type exhaustively, and also to formulate generalised
meanings that would, on a rather abstract level, represent all possible pragmatic
functions of a pattern. Such attempts have been made by e.g. Kohler (2006a) and
Niebuhr (2007b), which, however, do not cover all basic pattern types of German.

This study does not attempt to propose a new framework for the functional
description of the basic German patterns. The goal, as far as German is concerned,
is simply to elicit the basic patterns. For that, it was judged to be sufficient to
construct example contexts which lend a high probability for the basic patterns
to be produced by German speakers in the laboratory. However, it is not only
difficult to determine the functions, or usages, of the basic patterns, but also to
reverse the process, i.e. to define a situational context that would make the usage
of a certain pattern most likely. The general problem is that, as discussed in 3.2.3,
the choice of intonation pattern is not entirely determined by objective features
of the discourse, such as information states of discourse referents. These could
be controlled for. But the speaker still has the final choice of using a certain
intonation pattern in order to express some attitude, e.g. that s/he wishes to present
some information as newsworthy, although it is given in the discourse (cf. the
quotation by Baumann 2006 in 3.2.3).

To cope with this problem, several features were included in the test material
in order to increase the possibility of really obtaining the hypothesised patterns, as
discussed below. Short descriptions of the nine contexts, or conditions, together
with the expected intonation pattern for German, and a label which is used to refer
to these conditions, are listed in Table 5.2. The purpose of the brief descriptions
and the labels in Table 5.2 is to provide a short-cut to the full versions of the
constructed contexts, which are presented in Chapter 4.

In general, the choice of test contexts for this study was driven by the hy-
potheses derived from the literature (cf. 3.2).[2] However, for most of the pattern
types, several possible contexts, nine in total, are included in the material. For
instance, as discussed in 3.2.3, the (functional and formal) distinction between the
*medial* and the *late fall* is not as clear as between the *medial* and the *early fall*.
Therefore, four contexts in total were constructed for the elicitation of *medial* or
*late fall*s, mainly by including different nuances of *contrast*, since the *late fall* has
been associated with notions such as contrast, unexpectedness, or expressiveness.

---

[2]However, numerous informal pilot tests using different earlier versions of the test material
were performed, usually with 1–2 other native German speakers, which have finally resulted in the
material presented here.

Table 5.2: The nine test conditions and expected nuclear intonation patterns for German. See Table 5.1 for tonal labels of the pattern types.

| Label | Brief description | Expected nuclear pattern in German |
|---|---|---|
| NEWINFO | new-information response | *medial fall* |
| CORRECT | corrective response | *medial/ late fall* |
| EXCLAM | exclamation | *medial/ late fall* |
| SURPRIS | surprised feedback | *late fall* |
| CONFIRM | confirmative response | *early fall* |
| SUMMARY | concluding summary | *early fall* |
| QREPET | repetition question | *simple rise/ fall-rise* |
| QDISBEL | disbelieving question | *simple rise/ fall-rise* |
| ADDREQ | addressing for request | *fall-rise* |

Thus, the condition NEWINFO should rather clearly correspond to the *medial fall*, since it represents the signalling of new information without involving a specific contrast. The three conditions CORRECT, EXCLAM, and SURPRIS, all include a contrast of some sort. In CORRECT, the contrast is between the knowledge of a speaker B and an alternative proposed earlier by a person A, which is rejected by speaker B. In EXCLAM, there is a contrast between a speaker's own earlier belief and a new insight. Finally, in SURPRIS, the new information is in conflict with an expectation of the speaker.

The *early fall* can be expected to occur in utterances which in some way express established information, as represented in the material by CONFIRM and SUMMARY (cf. 3.2.3). Finally, *simple rise*s or *fall-rise*s can be expected in question-like utterances, and the *fall-rise* is also hypothesised to be the most likely pattern used when addressing a person in order to make a request, although admittedly, the usage of this pattern in this context has often been associated with child-to-adult speech in the literature.

However, an attempt was made to increase the possibility of eliciting the basic pattens not only by means of the quantity of test contexts, but also by means of qualitative differences between the contexts. As presented in 4.2 in more detail, two different elicitation techniques have been applied for two subsets of the material, which correspond to different contributions the test phrase would make to discourse (e.g. Linell and Marková 1993). In particular, the three conditions referred to as *responses* in Table 5.2 (NEWINFO, CORRECT, CONFIRM) represent answers to questions, which in the elicitation have been presented auditively. Hence,

NEWINFO, CORRECT, and CONFIRM may be classified as *locally-tied responses* in terms of Linell and Marková (1993).

In contrast, the six remaining conditions were elicited by means of a context description only, without a preceding context question. Three of them (QREPET, QDISBEL, ADDREQ) represent *initiations* rather than responses (Linell and Marková 1993)[3]. Condition SURPRIS is referred to as *feedback* (e.g. Sinclair and Coulthard 1975) in Table 5.2, but would be classified as a locally-tied response (like NEW-INFO, CORRECT, CONFIRM above) by Linell and Marková (1993). The remaining two conditions, SUMMARY and EXCLAM, are also primarily responsive in character, but they are not locally-tied; they respond globally to the communicative situation, rather than to the immediately preceding turn in discourse.

That is, the different conditions included for eliciting *early* (SUMMARY vs. CONFIRM) and *medial* or *late fall*s (NEWINFO, CORRECT vs. EXCLAM vs. SUR-PRIS) correspond to different *turn types* in terms of Linell and Marková (1993).

## 5.2   Method

This study is based on Corpus A. As explained in more detail in 4.2, 18 speakers in total, nine of each language, were recorded, and each subject produced the test phrase five times in each condition. The primary test phrase *in November* was used in eight conditions, while the secondary test phrase *Wallander* was used in one condition only (ADDREQ). In total, the corpus of this study (Corpus A) comprises 810 recorded utterances (9 conditions × 5 repetitions × 18 speakers).

The data analysis performed in this study consisted of an exploration and categorisation of the $F_0$ patterns obtained from the recorded utterances. For that, the recordings were inspected auditorily and visually using spectrograms and $F_0$ tracks in Praat (Boersma and Weenink URL). In a first, informal, step, general tendencies in the data were investigated. This inspection revealed that, first, all five basic patterns of German occurred in the corpus, and second, a majority of the intonation patterns produced both by the German and the Swedish speakers, can be classified as one of these patterns.

Hence, in a second step, a detailed auditory and visual examination of the individual recordings was undertaken in order to classify each utterance as belonging to one of the basic pattern types. However, the distinction between the *medial* and the *late fall* was not produced clearly in many cases. In a clear case of a *medial fall*, the stressed vowel is perceptually high-pitched (not rising), and is often realised with a more or less shallow rise in $F_0$. A clear case of a *late*

---

[3]The two conditions QREPET and QDISBEL have, of course, also a *responsive* character, but the *initiatory* aspects of these conditions seem to be primary.

Table 5.3: Criteria for the classification of $F_0$ patterns. The category *high+fall* comprises both patterns types *medial fall* and *late fall* from Table 5.1. Empty cells for the pre-stress syllable indicate that pre-stress pitch is irrelevant for the identity of the pattern.

| Pattern type | Label | Characteristic relative pitch features | | |
| --- | --- | --- | --- | --- |
| | | pre-stress | stress | post-stress |
| *early fall* | EF | high | low or falling | low |
| *high+fall* | HF | – | high or rising | low |
| *fall-rise* | FR | – | high or rising | falling-rising |
| *simple rise* | SR | – | low | rising |

*fall* would have a perceptually low stressed vowel and a high post-stress syllable, which can be realised by means of a rise in $F_0$ starting late in the stressed vowel and reaching an $F_0$ maximum in the post-stress syllable. However, in the material used in this study, with only one syllable following the stress, the rise is to be expected earlier, with a peak reached already in the stressed syllable (cf. 3.2.3). Hence, an $F_0$ rise in the stressed vowel is a typical correlate of both the *medial* and the *late fall*. A classification into *medial* and *late* would, of course, still be possible, but the distinction between these two patterns would rely almost exclusively on perceived pitch, while for all other distinctions, both pitch and (the visually inspected) $F_0$ could be used in the classification. Therefore, it was decided to subsume the *medial* and the *late fall* into a single category *high+fall* for the purpose of classification. That is, in the classification only four basic pattern types are distinguished, cf. Table 5.3.

Since the resulting four categories of pattern types (*high+fall*, *early fall*, *simple rise*, *fall-rise*) differ crucially in their characteristic pitch features (which are usually realised by corresponding $F_0$ features), the classification was generally trivial. The relevant pitch features used for the classification are listed in Table 5.3. These pitch features are approximate and relative. For instance, *low* in the final (post-stress) syllable can be realised as a fall in $F_0$, and can also be perceived as slightly falling, but relevant for the pattern type is only that pitch is perceived as low on that syllable in relation to a high or rising pitch on a syllable earlier in the utterance. Only a few cases could not be assigned to one of the four categories and were simply classified as *other*.

By using this simplified classification scheme, a reliable classification of pattern types can be guaranteed. At the same time, the potential distinction between the *medial* and the *late fall* is not entirely lost, since in a third step of analysis, the

realisation of the pattern types was examined separately for the test conditions. For that, $F_0$ was normalised using the method described in 4.2.3, and mean $F_0$ curves were plotted for the predominant patterns types of each test condition.

The time normalisation method requires a segmentation of the speech material. For the present study, the recorded utterances were segmented into six (test phrase *in November*) or five (test phrase *Wallander*) phonetic units, as illustrated by the broad phonetic transcriptions in (1), which represent a possible Swedish realisation. The German versions were segmented correspondingly.

(1)     *i november:* [(i)n] [ʊ]     [v] [ɛ] [mb] [bə(ɹ)]
        *Wallander:*        [(ʋ)a] [l] [a] [nd] [də(ɹ)]

As the transcriptions show, the initial function word *i* (Swedish) (or *im* for German) and the initial /n/ of *november* were subsumed in a single segment. One reason for this decision was that the $F_0$ pattern of the initial function word is less relevant for the investigation; another reason was that the preposition was often strongly reduced and occasionally elided, so that it would not have been possible to segment it in all recordings. Also, the initial /v/ in *Wallander* and the following vowel were subsumed in one segment. More precisely, the /v/ was typically realised as an approximant and hence acoustically manifested as an initial formant transition, which is difficult to separate from the vowel in a reliable manner; hence this formant transition was included in the vowel segment. Initial fricative portions, which also occurred occasionally, were excluded from the initial segment.

In general, segment boundaries were set at the salient acoustic landmarks typically occurring in connection with the vowels and consonants of the test material. For instance, [v] in *November* and the [l] in *Wallander* are typically characterised by an energy drop, which is clearly visible in the spectrogram. The boundary between the stressed and the post-stress syllable was set at the plosive burst for /b/, or /d/, respectively, because the boundary between a nasal and the occlusion phase of a following voiced plosive at the same place of articulation is usually a less salient landmark in the spectrogram. That is, the occlusion phase of the plosive was included in the preceding nasal segment.

## 5.3   Results

### 5.3.1   Distribution of pattern types

The distribution of pattern types that have occurred in corpus A, according to the classification described in 5.2, is displayed in Tables 5.4 (German speakers) and 5.5 (Swedish speakers). The tables show that only 10 utterances for German

Table 5.4: Relative frequencies (%) of $F_0$ patterns produced by **German** speakers. For $F_0$ pattern labels see Table 5.3 and Figure 5.1. Superscript: absolute frequency; number of speakers who produced the pattern in at least one of five repetitions. Example: $\mathbf{88.9}^{40;9}$ = pattern produced at least once by all 9 speakers, in 40 of 45 cases (88.9%) in total. Relative frequencies > 15% are printed in bold face.

| Condition | $F_0$ pattern type | | | | |
| --- | --- | --- | --- | --- | --- |
| | early fall | high+fall | simple rise | fall-rise | other |
| NEWINFO | $\mathbf{15.6}^{7;4}$ | $\mathbf{84.4}^{38;9}$ | | | |
| CORRECT | $11.1^{5;2}$ | $\mathbf{88.9}^{40;9}$ | | | |
| EXCLAM | | $\mathbf{82.2}^{37;8}$ | | | $\mathbf{17.8}^{8;2}$ |
| SURPRIS | | $\mathbf{91.1}^{41;9}$ | | $8.9^{4;1}$ | |
| SUMMARY | $\mathbf{93.3}^{42;9}$ | $6.7^{3;3}$ | | | |
| CONFIRM | $\mathbf{68.9}^{31;8}$ | $\mathbf{31.1}^{14;7}$ | | | |
| QREPET | | $4.4^{2;1}$ | $\mathbf{91.1}^{41;9}$ | $4.4^{2;1}$ | |
| QDISBEL | | $8.9^{4;1}$ | $\mathbf{66.7}^{30;8}$ | $\mathbf{20.0}^{9;3}$ | $4.4^{2;1}$ |
| ADDREQ | | $2.2^{1;1}$ | $\mathbf{51.1}^{23;6}$ | $\mathbf{46.7}^{21;6}$ | |

and 2 for Swedish were classified as *other*. The German cases are instances of a stress shift, or *contrastive stress* in terms of Bolinger (1961), where speakers have pronounced the test phrase with an accent on the (canonically) pre-stress syllable (*im NOvember*).[4] In the two Swedish cases, the test phrase was not produced with a nuclear pattern at all. These cases were instances of the condition ADDREQ, were the test phrase *Wallander?* was used. In the two problematic cases, the test phrase and the following question (*Skulle jag kunna få be dig om en tjänst?*) were produced as a single intonation phrase, and *Wallander?* was produced with a level pitch and judged to be unaccented.

Tables 5.4 and 5.5 show that the distribution of pattern types over the test conditions is in some respects similar for the German and Swedish data, but there are differences, as well. The *high+fall* pattern was the most frequently used pattern in the conditions NEWINFO, CORRECT, EXCLAM, and SURPRIS in both languages.

---

[4]Stress shift of this kind is not uncommon in German in contrastive utterances. However, stress shift was not expected to occur in this corpus. In my own native intuition a stress-shift in the word *November* would only occur if it were contrasted against another month with a name that also contains the string *-vember*, like *Okvember* or *Sepvember*. Thus, the only months where I frequently observe a stress-shift in my own and other's speech, are *JUni* and *JUli*, contrastively pronounced as *JuNO* and *JuLEI* with an additional change in vowel quality.

Table 5.5: Relative frequencies (%) of $F_0$ patterns produced by **Swedish** speakers. For $F_0$ pattern labels see Table 5.3 and Figure 5.2. Superscript: absolute frequency; number of speakers who produced the pattern in at least one of five repetitions. Example: $\mathbf{64.4}^{29;7}$ = pattern produced at least once by 7 of the 9 speakers, in 29 of 45 cases (64.4%) in total. Relative frequencies > 15% are printed in bold face.

| Condition | $F_0$ pattern type | | | | |
|---|---|---|---|---|---|
| | *early fall* | *high+fall* | *simple rise* | *fall-rise* | *other* |
| NEWINFO | $2.2^{1;1}$ | $\mathbf{93.3}^{42;9}$ | $4.4^{2;1}$ | | |
| CORRECT | | $\mathbf{100.0}^{45;9}$ | | | |
| EXCLAM | | $\mathbf{100.0}^{45;9}$ | | | |
| SURPRIS | | $\mathbf{100.0}^{45;9}$ | | | |
| SUMMARY | $\mathbf{28.9}^{13;3}$ | $\mathbf{66.7}^{30;8}$ | $4.4^{2;1}$ | | |
| CONFIRM | $\mathbf{28.9}^{13;5}$ | $\mathbf{64.4}^{29;7}$ | $6.7^{3;2}$ | | |
| QREPET | | $\mathbf{93.3}^{42;9}$ | | $6.7^{3;2}$ | |
| QDISBEL | $\mathbf{46.7}^{21;5}$ | $\mathbf{53.3}^{24;6}$ | | | |
| ADDREQ | $8.9^{4;2}$ | $\mathbf{31.1}^{14;5}$ | $\mathbf{20.0}^{9;2}$ | $\mathbf{35.6}^{16;4}$ | $4.4^{2;1}$ |

However, it was used more generally by the Swedish than by the German group. In the Swedish data, the *high+fall* occurred in at least 31% of the cases in all conditions; it was the most dominant pattern in all conditions except ADDREQ. In the German data, the *high+fall* also occurred in all conditions, but in SUMMARY, QREPET, QDISBEL, and ADDREQ, its occurrence can be counted as exceptional. Thus, the *high+fall* was the dominant pattern in four conditions only, as opposed to eight conditions for the Swedish data.

The *early fall* was the most usual pattern used by German speakers in SUMMARY and CONFIRM, although in CONFIRM, a *high+fall* also occurred in 31% of the cases; it also occasionally occurred in NEWINFO and CORRECT. As in the German data, the *early fall* was also used by Swedish speakers in SUMMARY and CONFIRM in a considerable proportion of the cases (29% each), but the dominant pattern for these conditions was still a *high+fall*. The *early fall* was uncommon in the Swedish versions of NEWINFO and CORRECT, but it was used in almost half of the cases in the condition QDISBEL, where it was never produced by a German speaker.

In the conditions QREPET, QDISBEL, and ADDREQ, the German speakers used predominantly a rising pattern. The *simple rise* was preferred in all three conditions, but also the *fall-rise* was used in almost half of the cases in ADDREQ, and

in 20% of the cases in QDISBEL. In the Swedish data, a rising pattern (*simple rise* or *fall-rise*) was preferred over a falling pattern only in the ADDREQ condition, where the *fall-rise* (36%) occured more frequently than the *simple rise* (20%), the latter being used systematically by two speakers only.

To summarise, the following tendencies in the distribution of pattern types were common in German and Swedish:

- A *high+fall* was the predominant pattern in NEWINFO, CORRECT, EX-CLAM, and SURPRIS.

- A concentration of *early fall*s occurred in SUMMARY and CONFIRM.

- Rising patterns (*simple rise* or *fall-rise*) were predominant in ADDREQ.

The most crucial differences can be summarised as follows:

- There is an asymmetry in the frequency of occurrence of the individual pattern types:

  - In the German data, all pattern types except the *fall-rise* occurred as the predominant pattern in at least two of the nine conditions.

  - In the Swedish data, a single pattern type (the *high+fall*) was predominant in all conditions except ADDREQ.

- There is an asymmetry in pattern distribution specifically for QREPET and QDISBEL:

  - In QDISBEL, the *early fall* was common in the Swedish data, but absent in the German data.

  - In QREPET and QDISBEL, the *simple rise* was common in the German data, but absent in the Swedish data.

## 5.3.2 Realisation of pattern types

Since the distribution of patterns has shown that all of the four basic pattern types have been used by Swedish and German speakers, and most of them also in several test conditions, the main purpose of this section is to examine differences and similarities in the productions of pattern types in the two languages and the different test conditions. The first, and most important part of this section attempts to pinpoint the general features of the pattern types as produced by German and Swedish speakers in the nine conditions. In a second part, more general differences between the German and Swedish productions are mentioned. A third part

briefly accounts for inter- and intra-speaker variation in the data, and finally, a fourth part comments on further phonetic parameters beyond $F_0$. It can be anticipated that the amount of variation found in the data does not have a severe impact on the characterisation of the general features provided in the first part of this section.

**General features of pattern types in the nine conditions**

Figures 5.1 (German) and 5.2 (Swedish) display the normalised mean $F_0$ curves of the dominant[5] $F_0$ patterns, averaged across all occurrences (by all speakers) of each pattern type in a given condition. For each language, the curves are grouped into the following four plots; the condition NEWINFO is included in all plots as a reference:

**(a)** CORRECT, EXCLAM, SURPRIS, NEWINFO

**(b)** SUMMARY, CONFIRM, NEWINFO

**(c)** QREPET, QDISBEL, NEWINFO

**(d)** ADDREQ, NEWINFO

As shown by plot (a) for both the German (Fig. 5.1a) and the Swedish (Fig. 5.2a) data, the range of the rising-falling $F_0$ movement in a *high+fall* is enlarged in SURPRIS and EXCLAM as compared to NEWINFO and CORRECT. Furthermore, in the German data, the range is on average larger in CORRECT than in NEWINFO, while the difference between SURPRIS and EXCLAM appears to be marginal. In the Swedish data, the difference in range between the conditions SURPRIS and EXCLAM is more pronounced than in the German data, while the range of the *high+fall* in CORRECT hardly differs from that in NEWINFO; the most pronounced difference between the Swedish curves for CORRECT and NEWINFO is rather a question of register, the curve for CORRECT being slightly higher, first and foremost in the stressed vowel.

The most deviating pattern among all versions of the *high+fall* in the (a) plots of Figures 5.1 and 5.2 is probably the German pattern produced in NEWINFO. Here, although the $F_0$ maximum is not reached earlier than at the onset of the post-vocalic /m/, the $F_0$ peak is rather broad, with a relatively high level already at vowel onset and a shallow rise through the vowel. This is to be compared with the

---

[5]That is, exceptional usages of pattern types are disregarded in the graphs for the sake of readability. A pattern type was counted as exceptional if it occurred in less than 7 cases (or less than 15%), which is an arbitrary decision. The cases which are included in the graphs are printed in bold face in Tables 5.4 and 5.5.

pattern for SURPRIS, where $F_0$ at vowel onset is slightly lower than for NEWINFO, while at the maximum, $F_0$ is higher in SURPRIS than in NEWINFO. In addition, the maximum in SURPRIS is reached considerably later than in NEWINFO. These characteristics in peak shape and timing reflect approximately what is to be expected for the German distinction between a *medial* and a *late fall* (cf. 3.2.3 and Niebuhr 2007b).

A corresponding distinction between a *medial* and a *late fall* is, however, not indicated by the average patterns for Swedish in panel (a). While SURPRIS and EXCLAM were produced with a wider peak range than NEWINFO, as in German, there was no corresponding similarity in the timing of the peak. On the contrary, the peak in the Swedish SURPRIS and EXCLAM productions is reached almost precisely at the boundary between vowel and post-vocalic /m/. That is, if we take the peak timing in NEWINFO as a baseline, which is, on average, reached slightly after the onset of the post-vocalic /m/ in both the Swedish and the German data, then the peak for SURPRIS and EXCLAM is reached considerably later than in NEWINFO in German, but slightly earlier than in NEWINFO in Swedish.

A further difference between the German and the Swedish productions of the *high+fall*, as observable in plots (a) in Figures 5.1 and 5.2, concerns $F_0$ on the pre-stress syllable (*no-* in *november*). In the German productions (Figure 5.1a), $F_0$ is rather flat on this syllable and on average equal to the $F_0$ value at the onset of the stressed vowel.[6] In the Swedish productions (Figure 5.2a, but also c), a slight rising-falling $F_0$ movement is observable on the pre-stress syllable, above all in the conditions EXCLAM, SURPRIS, and QDISBEL. The rising movements amounts to approximately 1–2 semitones, while the extent of the fall is more difficult to determine, because of the microprosodic $F_0$ dip in /v/. That is, there is a tendency towards a two-peaked pattern in the Swedish versions of the *high+fall*, as opposed to a one-peaked pattern in the German data.

Figures 5.1b and 5.2b display the curves for the most dominant patterns produced in the contexts SUMMARY and CONFIRM, where both *early fall*s and *high-+fall*s were produced. The *high+fall*s produced for CONFIRM (German and Swedish) and SUMMARY (Swedish only) share the general feature that the peak range is reduced as compared to the *high+fall* in NEWINFO. In the case of CONFIRM, there is, on average, hardly any rise in the stressed syllable as compared to the pre-stress syllable. The perceptual impression is nevertheless a relatively high-pitch stressed syllable, since the major final fall occurs in the post-stress syllable. These instances of the *high+fall* are probably better treated as *medial fall*s as found for NEWINFO in the German data.

---

[6]The dip in $F_0$ prior to the stressed vowel is a microprosodic effect due to the production of /v/, cf. 2.1.1.

Figure 5.1: Mean $F_0$ contours ($n$ in parentheses) of the most commonly produced patterns (the cases printed in bold face in Tab. 5.4) by **German** speakers. Pattern labels: HF = *high+fall*, EF = *early fall*, SR = *simple rise*, FR = *fall-rise*, cf. also Table 5.3. Time (x-axis) is normalised; the scale indicates the number of measurements; breaks in the curves and vertical lines symbolise segment boundaries (cf. 5.2 for segmentation criteria). Semitones (y-axis) relate to an approximation of the individual speakers' base $F_0$ (cf. 4.2.3).
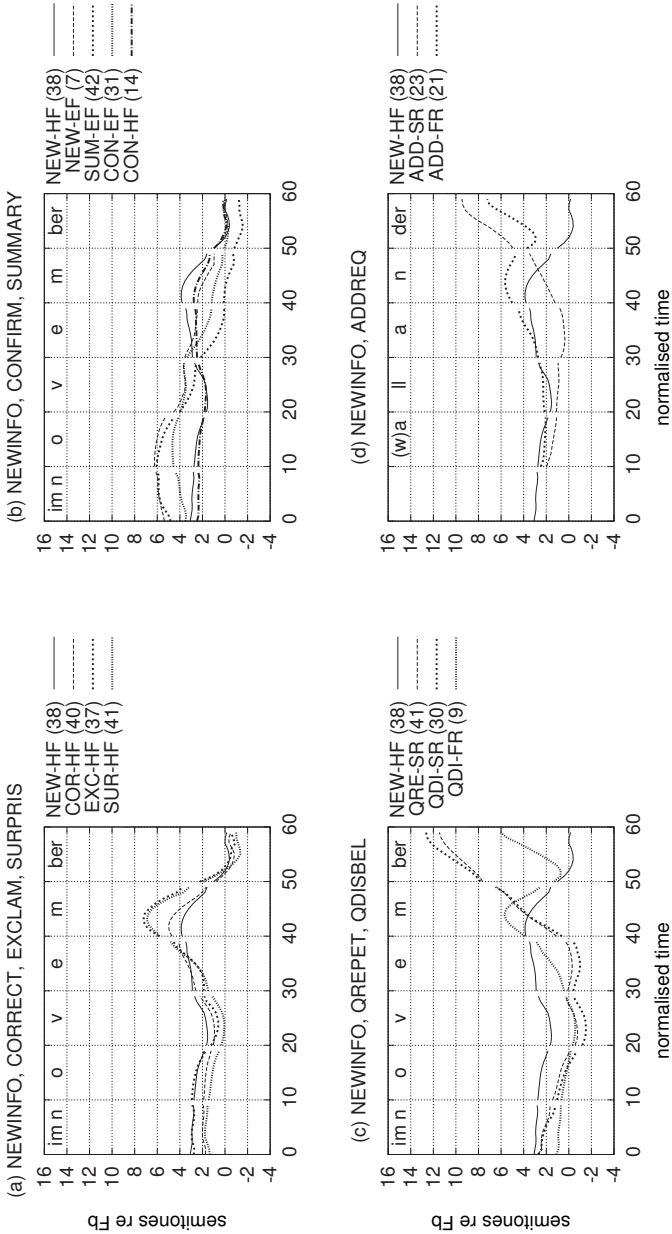
Figure 5.2: Mean $F_0$ contours ($n$ in parentheses) of the most commonly produced patterns (the cases printed in bold face in Tab. 5.5) by **Swedish** speakers. Pattern labels: HF = *high+fall*, EF = *early fall*, SR = *simple rise*, FR = *fall-rise*, cf. also Table 5.3. Time (x-axis) is normalised; the scale indicates the number of measurements; breaks in the curves and vertical lines symbolise segment boundaries (cf. 5.2 for segmentation criteria). Semitones (y-axis) relate to an approximation of the individual speakers' base $F_0$ (cf. 4.2.3).
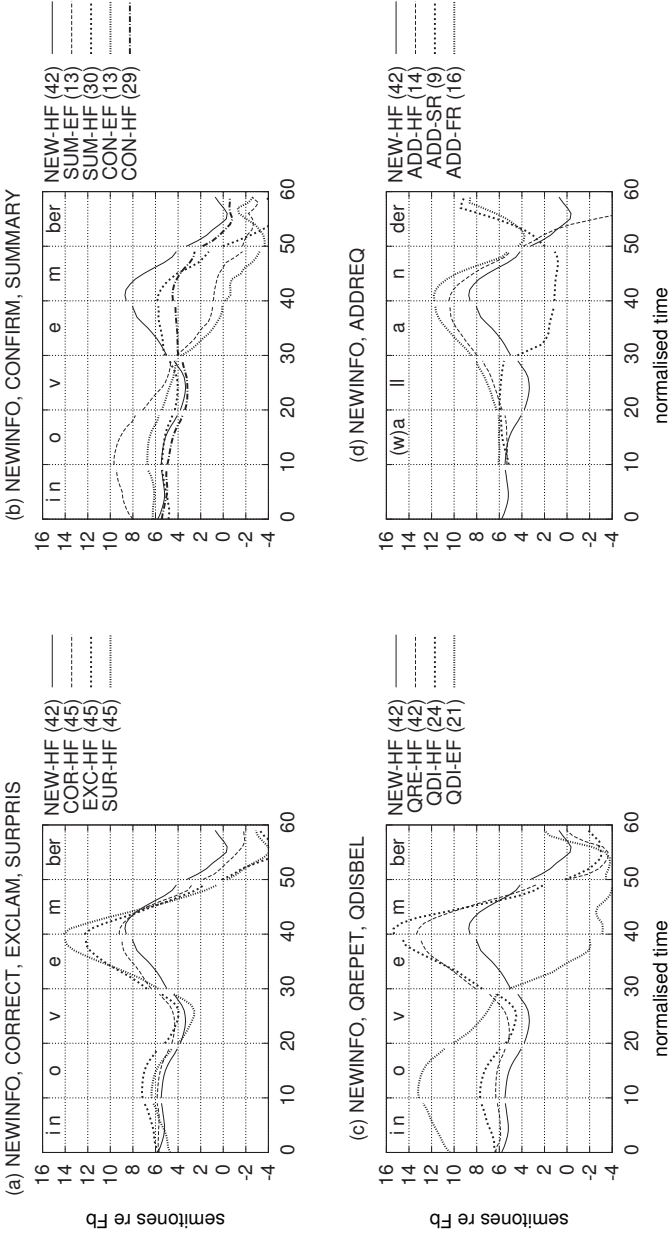
The (b) plots also display the curves for the *early fall*s that occurred in SUM-MARY and CONFIRM (and NEWINFO in the case of German). For both languages, the pattern was more pronounced in SUMMARY than in CONFIRM, reflected in a steeper fall from the pre-stress to the stressed syllable. In the Swedish data, an *early fall* occured also in condition QDISBEL (plot c). In this condition, the fall is even more pronounced than in SUMMARY.

Further versions of the *high+fall* were produced in QREPET, QDISBEL, and ADDREQ by Swedish speakers, as displayed in Figure 5.2 (c) and (d). The patterns for QREPET, QDISBEL, and ADDREQ are similar to the patterns produced in EXCLAM and SURPRIS since all are characterised by a higher $F_0$ maximum as compared to the pattern for NEWINFO. However, a difference in the domain of the $F_0$ range extension between QREPET, QDISBEL, and ADDREQ (plot c,d) on the one hand, and EXCLAM and SURPRIS (plot a) on the other hand is indicated by the data. In the latter two cases (plot a), it is first and foremost the $F_0$ movement in the vowel and the following consonant of the stressed syllable (*-em-* in *november*) that is expanded, while in the curves for the former three cases (panel c,d), the $F_0$ expansion starts earlier. This difference between a temporally restricted vs. a temporally wider expansion may also be characterised in terms of the $F_0$ at vowel onset, which is at a similar level in NEWINFO and SURPRIS (Figure 5.2a), but higher than NEWINFO in QREPET and QDISBEL (Figure 5.2c), and, most clearly, in ADDREQ (Figure 5.2d). As discussed below, these two strategies of $F_0$ expansion are not applied by all individual speakers; a clear example of the usage of the two strategies is speaker SM2 in Figure 5.5b.

Panels (c) and (d) also include the German and Swedish versions of *simple rise*s and *fall-rise*s. The shape and timing of the German *fall-rise* pattern produced in QDISBEL is similar to the German version of the *late fall* produced in SURPRIS and EXCLAM (panel a), with the simple distinction that $F_0$ in the final syllable is rising in the *fall-rise*, but falling in the *late fall*. The final $F_0$ value reached in the German *fall-rise* is, on average, similar to the $F_0$ value of the $F_0$ peak maximum in the stressed syllable. However, the *fall-rise* pattern produced by Germans in the ADDREQ condition differs from the *fall-rise* in QDISBEL. In ADDREQ, the pattern starts at a slightly higher level, and the rise in the stressed syllable is more compressed as in QDISBEL, while the final $F_0$ value is relatively high (i.e. higher than the peak maximum) in ADDREQ. The Swedish version of the *fall-rise*, as produced in ADDREQ, has basically the same shape as the *high+fall* as produced in ADDREQ, however, with a rise in the final syllable. Compared to the Swedish *high+fall* pattern in SURPRIS and EXCLAM, the peak of the patterns produced in ADDREQ (both the *high+fall* and the *fall-rise*) is compressed. In this respect, the *fall-rise* in ADDREQ is similar in the German and the Swedish versions. However, a difference between the German and the Swedish version is the relative height of the $F_0$ level reached in the final rise, which is higher than the peak maximum in

German, but lower than the peak maximum in the Swedish data.

Finally, the mean curves of the German *simple rise*s do not indicate any crucial systematic differences between the conditions QDISBEL, QREPET, and ADDREQ, except that in ADDREQ, the entire pattern is slightly compressed, as indicated most clearly by a lower final $F_0$ level in ADDREQ as compared to QDISBEL and QREPET. In all three conditions, the German *simple rise* is produced with a slight fall in $F_0$ from the pre-stress to the stressed syllable, followed by the final rise which starts in the stressed vowel. The Swedish version of the *simple rise*, produced by two speakers and in ADDREQ only, differs in shape from the German version. First, the onset of the pattern (the pre-stress syllable) is produced at a higher $F_0$ level and the step down to the stress is steeper and larger in the Swedish than in the German. Second, the remainder of the stressed syllable is produced at a low $F_0$ level, resulting in later onset of the final rise in the Swedish data, which comprises only the final syllable.

**Variations due to language and sex**

The Swedish speakers used a creaky voice quality in utterance-final syllables much more regularly than German speakers. This is reflected in Figures 5.1 and 5.2, which show that, in the case of falling patterns in German, most curves reach a final $F_0$ value around 0 on the semitone scale. This is to be expected considering the method of $F_0$ normalisation applied (cf. 4.2.3). However, most of the Swedish curves reach values considerably lower than 0, and there are also more irregular fluctuations in the final 1–2 segments of the Swedish curves. The $F_0$ patterns observed in the final syllables for falling $F_0$ patterns should hence be interpreted with care in the Swedish data.

A further difference between the German and the Swedish productions concerns the global $F_0$ range, which is, on average, wider for the group of Swedish than for the group of German speakers. This is exemplified in Figures 5.3a and 5.3b, where a selection of the curves presented in Figures 5.1 and 5.2 is rearranged showing both German and Swedish versions of the same pattern type in a single plot.

However, there appears to be an interaction in the data between language and sex, as suggested by Figure 5.3c. The figure displays the mean $F_0$ curves obtained for SURPRIS, separately for female and male speakers of each language. While the difference between the German and the Swedish version of the *high+fall* in SUR-PRIS seems large according to Figure 5.3a, a comparison with Figure 5.3c shows that the difference is mainly due to the productions of the German *male* speakers. The condition SURPRIS has been chosen to illustrate this interaction of sex and language because almost all speakers of both languages have produced the same pattern type in this condition (cf. Tables 5.4 and 5.5). The SURPRIS condition at-

(a) German vs. Swedish: NEWINFO, SURPRIS

(b) German vs. Swedish: SUMMARY, CONFIRM

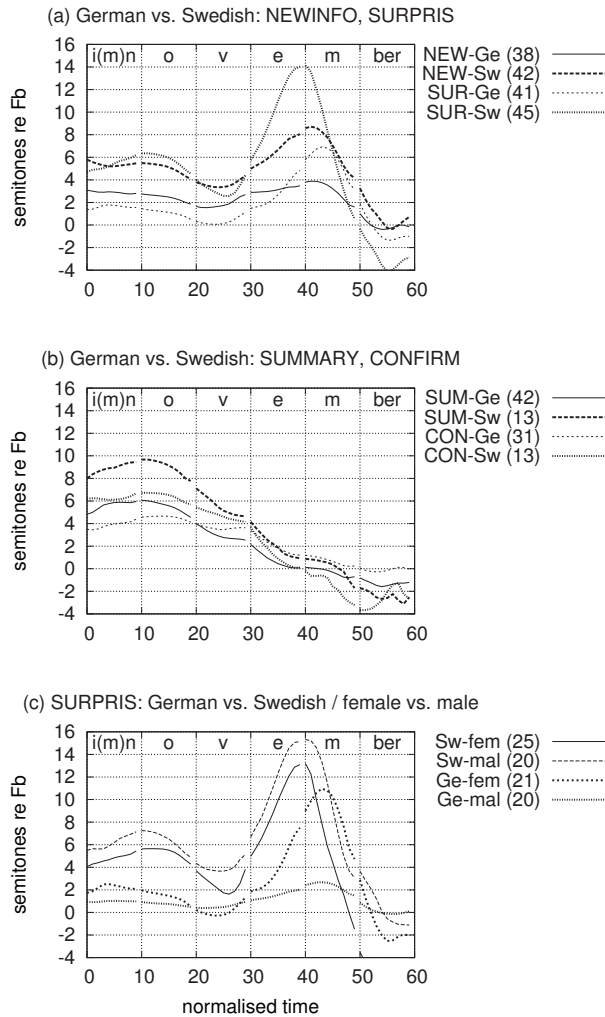(c) SURPRIS: German vs. Swedish / female vs. male

Figure 5.3: A comparison of **German** vs. **Swedish** productions. **(a)** *high+fall*s produced in NEWINFO and SURPRIS. **(b)** *early fall*s produced in SUMMARY and CONFIRM. **(c)** *high+fall*s produced in SURPRIS, separately for **female** and **male** speakers. Normalised mean $F_0$ curves ($n$ in parentheses; cf. Figures 5.1 and 5.2 for details).

tempts to elicit an expressive attitude. Hence, it could be the case that the German male speakers tested in this study were less involved or less comfortable with the task. However, the less expressive conditions show a similar interaction of language and speaker sex. It thus appears that the German male speakers included in this study deviate from both the female German and the male and female Swedish speakers regarding the $F_0$ range used in speech.

Figures 5.3a and 5.3c illustrate also more clearly the conclusion drawn above concerning the later timing of the $F_0$ maximum in the *high+fall* pattern for German as compared to Swedish speakers.

**Inter- and intra-speaker variation**

Besides the differences in pattern realisations between groups of speakers according to sex and language presented above, there is, of course, also between-speaker variation within these groups, as well as within-speaker variation.

Figure 5.4 provides an indication of different aspects of within-speaker variation found in the data. Each plot displays the $F_0$ curves of five repetitions produced by a single speaker in one specific condition. For the sake of illustration, the figure comprises one female and one male speaker of each language, and, in part, different conditions for the different speakers.

The upper two plots represent cases where all five repetitions were classified as the same pattern type. In panel (a) the $F_0$ curves are indeed almost identical, while in panel (b), there are at least four clearly different contours, which however, all fulfill the criteria defined for the pattern type *high+fall*. Both panels (a) and (b) represent rather extreme cases of minimal (a) and maximal (b) within-speaker variation within a given pattern type. In many cases, the variation is more moderate.

Figure 5.4 (c) and (d) illustrate another type of within-speaker variation. In both examples, the speaker has produced different pattern types among the five repetitions of a condition. In (c) one $F_0$ curve (*high+fall*) clearly deviates from the other four (*early fall*), while in (d) the speaker has produced two instances of a *high+fall*, and an *early fall* in three repetitions.

An example of between-speaker variation is provided in Figure 5.5. The figure compares realisations of the *high+fall* by two male speakers of Swedish. Both speakers have produced the *high+fall* in the five conditions NEWINFO, CORRECT, EXCLAM, SURPRIS, and QREPET, and the two panels of Figure 5.5 show, for one speaker each, the mean $F_0$ curve across the five (or four) repetitions of the *high+fall* for each condition.

The mean curves for speaker SM4 (Figure 5.5a) are quite close to the mean curves across all Swedish speakers (plots (a) and (c) in Figure 5.2). That is, the curves for NEWINFO and CORRECT on the one hand are similar, and differ clearly
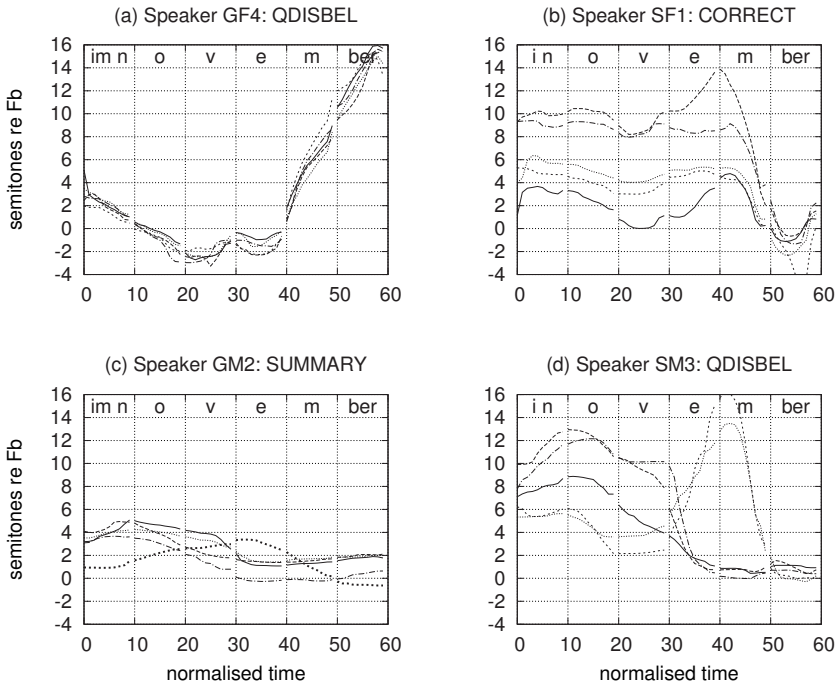
Figure 5.4: Examples of intra-speaker variation. Four speakers: female (top), male (bottom), German (left), Swedish (right); one example condition per speaker. Each plot shows normalised $F_0$ contours of all 5 repetitions (single cases, $n = 1$ per curve). See Figures 5.1 and 5.2 for details concerning the axes.

from the curves for EXCLAM, SURPRIS, and QREPET on the other hand, which are produced with a widened $F_0$ range in the stressed syllable. However, as discussed above, the mean curve across all Swedish speakers for QREPET (panel (c) of Figure 5.2) differs from the curves for EXCLAM and SURPRIS, in that the pattern in QREPET is not only characterised by a local increase in $F_0$ range restricted to the stressed syllable. Instead, in QREPET (and QDISBEL, ADDREQ), the expansion of the $F_0$ range sets in earlier. This distinction of different types of range increase is not observable in the data for SM4, who thus seems to have used only two types of *high+fall*s, which differ in peak height, while three types can be found for the group of Swedish speakers.

However, speaker SM2 has produced finer distinctions between the test conditions. Like SM4, SM2 also used an increased $F_0$ range in EXCLAM, SURPRIS, and QREPET as compared to NEWINFO, but SM2 seems to have distinguished be-
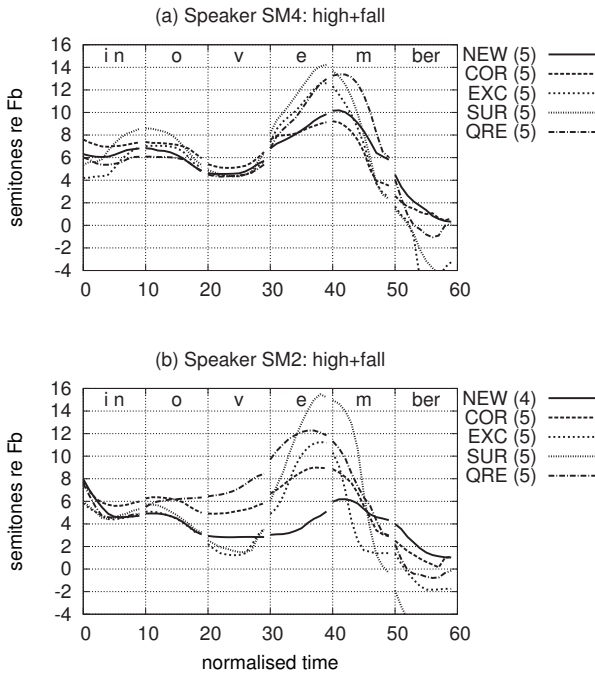
Figure 5.5: A comparison of *high+fall*s produced in NEWINFO, CORRECT, EX-CLAM, SURPRIS, and QREPET by two male speakers of Swedish. Normalised mean $F_0$ curves ($n$ in parentheses; cf. Figures 5.1 and 5.2 for details).

tween all five conditions, in part by means of different degrees of $F_0$ peak height, but also by means of two different strategies for the range increase, as discussed above for the group mean curves. That is, while EXCLAM and SURPRIS constitute two degrees of the temporally restricted type of widened $F_0$ range, CORRECT and QREPET constitute two degrees of the temporally broader type of increased $F_0$ range.

These examples illustrate that the mean curves across all speakers (Figures 5.1 and 5.2) do not represent all individual speakers perfectly, but there are no dramatic deviations of individual realisations from the patterns obtained for the groups of speakers.

**A note on parameters beyond $F_0$**

The research question of this study is whether Swedish and German share a common inventory of basic nuclear intonation patterns. Since these basic patterns have been defined in very general terms, and since it is well known that $F_0$ is the primary acoustic correlate of pitch, which in turn is the primary psychoacoustic correlate of intonation, this study has been restricted to $F_0$ as a correlate of the basic intonation patterns. Moreover, by using time-normalised $F_0$ contours, all duration information has been eliminated from the analysis. However, in order to pinpoint the phonetic realisations of the test phrase in the different conditions, and hence to understand the phonetic correlates of the corresponding prosodic patterns more completely, it would, of course, be neccessary to include further phonetic parameters in the study.

Informally, it can be noted that in the present data, duration, voice quality, and articulatory effort vary greatly, and there might be some correlations of these parameters with the functions of the patterns. An example for variations in articulatory effort present in the data are variations in the realisation of the stress-initial /v/ in *November*, which is realised both as [ʋ] and [v] in the data, in the latter case with different degrees of voicing and fricative noise, resulting in [v̥] in some cases. A pronunciation typically sounds more emphatic, the more fricative noise and the less voicing there is in the production of /v/.

## 5.4 Discussion

The goal of this initial study was to investigate whether Swedish has a similar inventory of basic nuclear pattern as German. A set of five basic nuclear patterns of German was defined as the point of departure (Table 5.1). In the analysis of the distribution of pattern types over the nine conditions, as explained in 5.2, two of the five basic patterns (the *medial* and the *late fall*) were subsumed in a single category *high+fall* in order to provide reliable criteria for the classification of the data.

### 5.4.1 Summary and comments on the results

As far as German is concerned, all four (actually five, cf. below) pattern types have occurred in the data in approximately those conditions where they were expected according to the hypotheses (cf. Tables 5.2 and 5.4). The main deviations between the hypotheses (Table 5.2) and the results (Table 5.4) concern the usage of the *high+fall* in CONFIRM in 31% of the cases, and the usage of the *simple rise* in ADDREQ in 51% of the cases. However, in CONFIRM, the expected pattern

(the *early fall*) was still dominant, and the two alternative patterns occurring in ADDREQ (the *fall-rise* and the *simple rise*) at least share the common feature of a final rise.

The four pattern types (*high+fall*, *early fall*, *fall-rise*, *simple rise*) also occurred in the Swedish data. The results reveal some differences and similarities both concerning the usage and the realistion of these pattern types between German and Swedish. It must be kept in mind that the conclusions drawn here for Swedish are limited to accent I.

Common in the distribution of pattern types in the Swedish and the German data was first, that the *high+fall* was the most frequently used pattern in four conditions (NEWINFO, CORRECT, EXCLAM, SURPRIS). Second, in both languages, the *early fall* occurred frequently in CONFIRM and SUMMARY (although for Swedish, it was not the most frequent pattern in these conditions, cf. below). Third, a rising pattern (*fall-rise* or *simple rise*) was dominant in ADDREQ.[7]

That is, a high or rising pitch (*high+fall*) was used in the conditions where the speaker either used the accent to mark new information (NEWINFO, CORRECT) or a contrast of some sort (CORRECT, EXCLAM, SURPRIS). According to Gussenhoven (2002), this form–function relation can be explained by means of the *effort code* (cf. 2.4): The presence of a high-pitched accent indicates, in a metaphorical manner, a greater articulatory effort, which is to be interpreted in the way that the speaker wishes to signal the accented word as especially important, in the present cases either due to newness of the information or due to contrast. The effort code thus also provides an explanation for the variations in peak height (or expansion of the $F_0$ range) in the *high+fall*s produced in NEWINFO, CORRECT, EXCLAM, and SURPRIS, since a higher peak is used in more contrastive or expressive conditions.

However, different types of variation of the *high+fall* have been observed for German and Swedish. For German, the two hypothesised basic patterns *medial* vs. *late fall*, which were subsumed in the classification, could be disentangled in the analysis of the $F_0$ mean curves. The *medial fall* occurred in the NEWINFO condition and differs from the *late fall* (CORRECT, EXCLAM, SURPRIS) in the shape of the $F_0$ rise in the stressed syllable (shallow vs. steep), in $F_0$ peak height, and $F_0$ peak timing. Although a comparable variation of $F_0$ peak *height* has also been observed in the Swedish data, a corresponding *timing* distinction between a *medial* and a *late* fall has not been found for the Swedish speakers. On the contrary, in the conditions where a *late* fall was found in the German data, the $F_0$ peak tended to occur slightly earlier in the Swedish data as compared to the pattern produced in NEWINFO (cf. Figures 5.1a, 5.2a, 5.3a). A likely explanation for this

---

[7]A preliminary report on the results for the condition ADDREQ has been presented in Ambrazaitis (2008). The group of speakers included in Ambrazaitis (2008) overlaps with the present group of speakers, but the two groups are not identical. However, the results and conclusions are generally the same in the two reports.

phenomenon is a timing restriction related to the enhancement of the word accent contrast. The $F_0$ peak in the *high+fall* in the present data reflects the focal accent (H-) of the Lund model. In accent II simplex words, the H- peak is produced in the post-stress syllable, which implies a salient perceptual cue for accent II. Hence, if in an accent I word the H- peak is produced too late, this resulting pattern could be perceptually similar to accent II, despite the lack of typical accent II fall in the stressed syllable (H*+L).

Instead of a timing distinction found for German, two different strategies of expanding the $F_0$ range in the *high+fall* were found in the Swedish data. In the conditions EXCLAM and SURPRIS, the expansion was local, restricted to the stressed syllable. Unlike German speakers, the Swedish speakers also used the *high+fall* pattern in the question-like conditions QREPET, QDISBEL, and ADDREQ. In these conditions, the $F_0$ was expanded (as compared to NEWINFO) in a temporally less restricted manner. A possible interpretation of this result is that an extended *global* $F_0$ range is applied in QREPET, QDISBEL, and ADDREQ. Hence, the result seems to be in line with Gårding's (1979) account of question intonation in terms of a modification of the *tonal grid* (cf. 3.3.3). Another possible interpretation in AM terms would be to assume that an *upstep* of the low word accent tone (L* in accent I: H+L*) is applied for a certain communicative purpose, i.e. in the question-like conditions QREPET, QDISBEL, and ADDREQ in the present data. It is not attempted here to arrive at the most adequate description of this phenomenon. However, by contrasting the results for questions with the results for expressive or contrastive utterances like EXCLAM and SURPRIS, this study suggests that a raised peak, or an expanded peak range, may be due to two separate mechanisms. It seems possible to apply these two mechanisms simultaneously, as shown by the data for speaker SM2 in Figure 5.5. Further studies are required to investigate these two strategies and their usage in more detail.

Kohler (2006a) discusses the *high+fall* in connection with the *early fall* and explains the functional relation between the two pattern types by means of the *frequency code* (cf. 2.4; Ohala 1984). The *early fall* is typically used, as in this study, in connection with established information or the closing of a discussion. As mentioned in 3.2.3, Kohler (2006a) reduces the basic meanings of the *medial* and the *early* peak[8] to the basic pragmatic meanings *openness* vs. *finality*. Kohler (2006a) relates these meanings to *submissiveness* vs. *dominance*, which are the general interpretations of high vs. low pitch in the frequency code. While the explanation according to the *effort code* (Gussenhoven 2002) provided above accounts for the *presence* of an accent and its realisation, Kohler's (2006a) reference to the *frequency code* thus explains the variation of accent *type*.

---

[8]The term *peak* is used to refer to the *accent*; in this study, the terms *early*, *medial*, or *late fall* are used in order to refer to the entire *nuclear pattern*.

Also the choice between a final rise and a fall in questions has been related to the frequency code for German (Kohler 2005b) and Swedish (House 2005). The essence of both accounts is that a fall is used when the question is oriented towards the factual information, while the rise implies an orientation towards the addressee. However, a difference between German and Swedish, as reported in the literature, is that a final rise is much more general in German and does not necessarily imply a specific addressee-orientation (e.g. in *yes/no*-questions, cf. Kohler 2005b), while in Swedish, a rise has generally been treated as an optional feature of questions (Gårding 1979) until House (2005) suggested that the final rise actually has a specific function in social interaction. These earlier conclusions are in line with the present data, where ADDREQ most clearly represents a context where the speaker–listener interaction is at the centre of the communicative situation. This condition was the only one where a final rise occurred (as the most dominant pattern) in the Swedish data, while for German, a rise also occurred in QDISBEL and QREPET.

A difference between the German and Swedish realisations of the *simple rise* was the fall in $F_0$ from the pre-stress to the stressed syllable, which was found to be much more pronounced in the Swedish than in the German data. A possible explanation is that this fall in the Swedish data reflects the contribution of the word accent (accent I: H+L*). As discussed in 3.3.1, this fall is often reduced when the word is associated with an H-, but clearly pronounced otherwise. This is reflected in the present results, where all patterns of the type *high+fall* clearly represent instances of the H- accent, while the *simple rise* and the *early fall* do not contain an H- (cf. 5.4.2 below): The word-accent fall is most clearly produced in the *simple rise* and in the *early fall*.[9] Also in the *high+fall*s, an $F_0$ peak on the pre-stress syllable can be observed in the average data, but its extension is reduced as compared to e.g. the fall in the *simple rise* (cf. Figure 5.2).

A crucial difference concerning the usage of the pattern types in the present data is that the *high+fall* (representing the H- L% in the Lund model) has been used as a kind of default pattern in all but one (ADDREQ) of the conditions. That is, while the other pattern types have also occurred in several conditions, the *high+fall* was still more frequent. In contrast, in the German data, each of the four pattern types occurred as the most dominant pattern in at least two different conditions, as expected by the hypotheses. An extreme example, as mentioned above, is the case of QREPET and QDISBEL, where a *simple rise* was the typical pattern used by German speakers, while it never occurred in the Swedish data.

However, the difference in pattern distribution between the German and the Swedish data is not only related to a more restricted usage of the variety of pat-

---

[9]In the *early fall*, however, the pattern does not differ crucially from the German version. The treatment of the *early fall* is discussed in more detail below.

terns by Swedish speakers. Most unexpectedly, the *early fall* was frequently used by Swedish speakers in QDISBEL, where it was never used by German speakers. However, what unites the condition QDISBEL with CONFIRM and SUMMARY is that in all three conditions, the accented word (*november*) represents established information, which is *stated* in CONFIRM and SUMMARY, but *questioned* in QDISBEL.

To summarise, all four pattern types under investigation have been used by German and Swedish speakers, and some common form-function relations have been found, which can be related to biological codes (cf. 2.4; Ohala 1984; Gussenhoven 2002). Besides these similarities, German and Swedish (accent I) differ in the realisation of the pattern types, and more crucially in their usage. The most salient difference is that in Swedish, it seems to be adequate to use a single pattern (the *high+fall*) almost universally. This finding is well in line with the common modelling of Swedish intonation, and might indeed explain the fact that nuclear patterns other than this *high+fall* have hardly been recognised before. However, this study shows that there actually are further pattern types even in Swedish, and their functions are basically the same as in German. Hence, it might be concluded that Swedish indeed has a similar inventory of nuclear patterns as German, but their usage is more restricted.

### 5.4.2   Modelling of nuclear patterns in Swedish

The results of this exploratory study provide an initial evidence for the overall hypothesis, namely that Swedish has a similar inventory of intonation patterns as German, since four basic pattern types have occurred in the data for both languages. Table 5.1 shows that only two of these patterns are predicted by the Lund model: the *high+fall* (H- L%) and the *fall-rise* (H- LH%) (although the *fall-rise* in Swedish has so far not explicitly been associated with other functions than the signalling of continuation, cf. Gussenhoven 2004). The two additional patterns found in the data for Swedish are the *simple rise* and the *early fall*. As discussed in 3.3.4 and 5.4.1, the *simple rise* has already been reported for Swedish by House (2005).

What the two additional patterns (the *simple rise* and the *early fall*) have in common is a *low* pitch in connection with the stressed vowel, as opposed to a *high* pitch in the *high+fall* and the *fall-rise*. A straightforward way to account for the two additional patterns is to assume that Swedish not only has a *high* sentence accent (H-), but also a *low* one (which may preliminarily be represented as L-). The corresponding revised pattern inventory of Swedish is included in Table 5.6.

The table suggests that the proposed L- does not symbolise any additional melodic movement. Hence, it might be argued that the additional L- would represent a superfluous model category and should therefore be rejected by Occam's

Table 5.6: A revised comparison of basic nuclear intonation patterns, as originally presented in Table 5.1. The revision for Swedish includes an additional low (L-) sentence accent. The word accent pattern for accent I is included for Swedish.

|              | German   | Swedish (accent I) |
|--------------|----------|--------------------|
| *early fall*   | H+L* L%  | (H+L*) L- L%       |
| *medial fall*  | H* L%    | (H+L*) H- L%       |
| *late fall*    | L*+H L%  |                    |
| *fall-rise*    | H* LH%   | (H+L*) H- LH%      |
| *simple rise*  | L* LH%   | (H+L*) L- LH%      |

razor. Instead, the two additional patterns (the *early fall* and the *simple rise*) could also be accounted for by simply assuming a lack of the focal accent (H-). That is, they could be regarded as instances of a non-focal accent I (only accent I was tested in this study), combined either with the low or the rising boundary tone, represented as H+L* L% (*early fall*) or H+L* LH% (*simple rise*). This description of the present data is referred to as the 'lack of H-' account in the following discussion.

However, the main purpose of the proposed L- accent is to indicate a phrase-level prominence, which is to be distinguished from a word-level prominence. This interpretation of the data is to be preferred over the 'lack of H-' account, as argued in the remainder of this section.

The *early fall* and the *simple rise* occurred in this study as nuclear patterns in the sense that they contain the most prominent, and indeed, the only accent of the intonation phrase. It is obvious that they have functions which are related to the *utterance* (or the discourse). Hence the most intuitive description of these patterns, as proposed here, is in terms of an *utterance*-level accent (L-). Implicitly, a suggestion in this vein has been made already by House (2005), who treated the *simple rise* as a late-timed H- accent. That is, House (2005) did not explicitly suggest a *low* accent, but he recognised the accent involved in the nuclear *simple rise* as a phrase-level phenomenon.

However, the non-focal variants of the word accents assumed by the Lund model are intended to account for a crucially different phenomenon. The Lund model has been formulated based on speech material where each intonation phrase contains at least one H- accent, which co-occurs with a word accent gesture (e.g. Bruce 1977). Besides this focal accent there can be non-focal accents in the phrase, containing a word accent gesture only (H+L* for accent I; H*+L for accent II). That is, the non-focal word accents (H+L* and H*+L) in the Lund model

represent either pre-focal or post-focal accents. However, the nuclear accents in
the *early fall* and the *simple rise* in the present data are neither pre- nor post-focal.
The relation between sentence-level (focal) and word-level (non-focal) accents
has been modelled as a difference in phonological prominence level. While the
domain of the sentence accent (H-) is the prosodic phrase, the domain of the word
accents (H+L* or H*+L) is the prosodic word (cf. 2.1.2, and e.g. Bruce 1998).

Thus, the 'lack of H-' account would have serious theoretical implications.
Either it would have to be assumed that the utterances containing an *early fall*
or a *simple rise* in the present data do not contain any phrase-level accent at all,
although the nuclear pattern clearly has a sentence- or discourse-level function.
This assumption is not only counter-intuitive, but it would also imply a further
problem. Although the distribution of patterns was not the same for German and
Swedish, there are salient parallels concerning intonational form–function rela-
tions in the two languages (cf. 5.4.1). Most relevant for the present discussion
is that the Swedish *simple rise* occurred in a context where also German speak-
ers produced a *simple rise* (ADDREQ), and also for the *early fall* there are two
overlapping contexts (CONFIRM, SUMMARY). However, if the *early fall* and the
*simple rise* in Swedish were analysed as lacking a sentence accent, this would
imply crucially different underlying representations of German and Swedish nu-
clear patterns, despite their phonetic and functional similarity. From a typological
perspective, it would seem more appropriate to account for these *phonetic* and
*functional* similarities in the two closely related languages by also assuming a
*phonological* similarity of the German and the Swedish pattern.

Alternatively, the 'lack of H-' account would have to assume that 'word ac-
cents' have the same prominence domain as the sentence accent (H-), which would
imply a radical modification of the standard description of Swedish prominence
levels (cf. 2.1.2).

That is, while the 'lack of H-' account at a first sight seems to be a simpler way
of describing the present data, because it does not propose an additional model
category like the L-, it is unsatisfactory. It either fails to account for the present
data in a typologically plausible manner, or it implicitly introduces a radical mod-
ification of basic assumptions on Swedish prominence levels. In contrast, the
assumption of an additional low sentence accent in Swedish (L-) accounts for the
common form–function relations in German and Swedish while it at the same time
avoids the introduction of dramatic modifications of basic assumptions concern-
ing Swedish phonological modelling. Thus, Occam's razor would indeed suggest
the preference of the L- proposal over the 'lack of H-' account, rather than vice
versa.

It should be stressed that the essential feature of the L- proposal is that it is
necessary for a comprehensive understanding of Swedish intonation to recognise
that Swedish has a paradigmatic choice of sentence accents, like German. Besides

a high-pitch accent (H-) there is also a low-pitch accent. How this contrast is to be represented is, at least in part, a matter of the descriptive framework and not the primary issue. The preliminary representation of this low accent chosen here for the AM-framework is L-, but this choice has only been motivated by the analogy with H-.[10] Hence, it might well be the case that further consideration will arrive at another representation of this low accent.

For instance, an alternative interpretation of the data would be to assume a *falling*, instead of a *low* accent in Swedish, i.e. H+L- instead of L-. On the one hand, the assumption of an L- seems sufficient, since the observed falling $F_0$ movement in the *early fall* and in the *simple rise* can be explained by the H+L* of accent I. However, a falling accent (H+L-) would more adequately account for the observed form–function relations of the *early fall* vs. the *high+fall*, which are similar in Swedish and German. The H+L- would imply that the falling portion in the *early fall* may have a double nature: It is determined by utterance prosody on the one hand (as in German), but it would be predicted by the word accent anyway. It seems that the resolution of this issue is an empirical question: It is necessary to investigate how the L- (or H+L-) accent is realised in connection with accent II. This is the research question treated in Chapter 7.

### 5.4.3   Conclusion: the L- hypothesis

This study has shown that Swedish has two further nuclear patterns which are not recognised by the Lund model (or Swedish intonation models in general) yet: an *early fall* and a *simple rise*. It has been argued in 5.4.2, that both the *early fall* and the *simple rise* imply a low or a falling utterance-level accent (H+L-/ L-). So far, this conclusion has been based on theoretical arguments only, and should therefore be regarded as a hypothesis, as stated explicitly in (2).

(2)     **The L- hypothesis:**
        Besides the high phrase-level accent (H-), Swedish also has a low phrase-level accent (L-).

One purpose of the following studies presented in this thesis is to test this hypothesis empirically. For that, the studies concentrate on the *early fall*. Two lines of evidence are presented: First, it has been argued above that the *early fall*, like the *high+fall*, should be treated as a phrase-level phenomenon. This view would be convincingly supported if the *early fall*, like the *high+fall*, could be used to signal individual words in an utterance as more prominent than other words in the same utterance. In other words: Is there a phonetic difference between an H+L* L- and

---

[10]The discussion in this section has, hopefully, shown that the proposal of an L- sentence accent in Swedish is not merely made in order to fill a gap in the phonological model of Swedish.

a (pre- or post-focal) H+L*? This is tested in Chapter 6, still in connection with accent I only.

Second, Chapter 7 investigates the interaction of the nuclear pattern (*early fall* vs. *high+fall*) with the signalling of the word accent category (accent I vs. accent II). The results on the realisation of accent II in connection with an *early fall* lend further support for the hypothesis. Moreover, the final studies (Chapter 7) also treat the question as to whether the nuclear *early fall* in Swedish should be analysed as resulting from a low (L-) or a falling (H+L-) sentence accent.

# Chapter 6

# Focus, topic, and accent type in German and Swedish

## 6.1   Introduction

The previous chapter was concerned with German and Swedish nuclear intonation patterns in an utterance comprising a single prosodic word. These patterns were elicited by means of nine test conditions, which simulated different discourse functions of the test phrase. One finding was that, in both German and Swedish, the utterance 'in November' was produced with a *high+fall* (cf. Table 5.3) in conditions that are related to new information or some sort of contrast or expressiveness (NEWINFO, CORRECT, EXCLAM, SURPRIS). This *high+fall* pattern reflects an H* or an L*+H accent (plus final L%) in the GToBI description of German, and an H- (focal accent) in the Lund model description of Swedish. However, in conditions related to the confirmation of information or the closing of an argument (CONFIRM, SUMMARY), the *early fall* (cf. Table 5.3) was the most common pattern in the German productions, and also frequently used by Swedish speakers. This *early fall* represents an H+L* accent in German. For Swedish, the Lund model accounts for a phonetically similar, but functionally distinct pattern, namely the non-focal accent I (H+L*). It was argued that these two phenomena – a non-focal word accent gesture vs. a nuclear accent – should be distinguished in a model of Swedish prosody, since the nuclear *early fall* represents an utterance-level phenomenon also in Swedish, as in German. Therefore, the **L- hypothesis** was proposed. The study presented in this chapter is based on Corpus B and undertakes a first attempt to test the L- hypothesis empirically.

In 3.2.3, a basic inventory of six typical German intonation patterns was presented, which represent the point of departure for the comparison of German and Swedish intonation in this thesis. Five of these patterns are nuclear patterns, which

consist of one accent and a final boundary pattern. These have been treated in chapter 5, which was based on Corpus A. The sixth pattern type presented in 3.2.3 is the *hat pattern*, which comprises two accents (a nuclear and a prenuclear one), connected by a high plateau. One typical context of a hat pattern in German is an information structure consisting of a *contrastive topic* followed by a focus. This structure was elicited in Corpus B, where a three-word sentence was used (as opposed to the one-word phrase in Corpus A), and is treated in the present chapter. The intonational marking of a *contrastive topic* is thus a secondary object of study in this chapter.

### 6.1.1   Focus usage and accent type

The L- hypothesis assumes that the *early fall* found as a nuclear pattern, e.g. in confirmations, is an utterance-level accent, like the *high+fall*. If it could be shown that the *early fall*, like the *high+fall*, can be used to signal individual words in an utterance as more prominent than other words in the same utterance, this would imply convincing evidence for the L- hypothesis. However, the design of Corpus A, used in chapter 5, cannot provide such evidence, since the test phrase only contains a single prosodic word ('in November'). Therefore, in Corpus B, used in the present chapter, a test sentence containing three content words was used.

   One reason for highlighting individual words at the utterance level is to signal a *narrow focus* (cf. 2.3.1). It has been discussed in 2.3.1 that narrow focus signalling obviously involves accent *placement*, but that also the *type* of accent may be relevant for the signalling of focus in different *pragmatic usages* (Krifka 2007), e.g. in *confirmative* as opposed to *corrective* utterances, or simply when new information is provided in reply to a *wh*-question, a case which can be referred to as a *'plain' focus*. The test condition NEWINFO from chapter 5 represents such a *plain* focus, while the condition CONFIRM may be said to represent a *confirmative* (usage of) focus. However, in chapter 5 only the paradigmatic aspect of focus signalling (i.e. type of accent)was considered.

   The present chapter concentrates on two of the conditions from chapter 5 – new-information responses (NEWINFO) and confirmative responses (CONFIRM) – and takes into account a further pragmatic dimension, namely *narrow focus location*. The conditions NEWINFO and CONFIRM were chosen for this study, because they are easily elicitable by means of context questions, a property which is regarded important for the simultaneous elicitation of focus location.

   If the falling pattern found in a confirmation (H+L* L-) includes an utterance-level accent (L-), then it should differ formally and functionally from a similar falling pattern (H+L*) which, however, does not represent an utterance-level accent, but a (non-nuclear) word accent. The specific hypothesis is that the rising-falling nuclear pattern ((H+)L* H- L%) of Swedish is used to signal *narrow focus*

rendering *new information* (*plain* focus), while the falling pattern (H+L* L- L%) can be used to signal *narrow focus* rendering *given information* (e.g. in a *confirmative* focus). In such confirmative cases it is thus expected to find utterances lacking any H- accent, instead exhibiting a *falling pattern* on each word. However, according to the hypothesis one of these falls should function as a *'focal fall'* (H+L* L-), and for that, it should differ phonetically – i.e. acoustically and perceptually – from the (pre- or post-focal) falls (H+L*).

However, one should keep in mind that the L- hypothesis actually only assumes that a nuclear *early fall* represents an utterance-level prominence. It does not directly predict that this prominence can be used to signal narrow focus in a confirmative sentence containing more than one prosodic word. That is, it might still be the case that a Swedish speaker normally would avoid producing confirmations using a non-elliptical sentence like the test sentence of Corpus B, and even more signalling a specific word in such a confirmation as narrowly focussed. However, if it can be shown that an early fall indeed can signal focus in such a context, then the L- hypothesis is strongly supported.

The data analysis comprises two parts. First, $F_0$ patterns are analysed for the German and the Swedish data. Motivated by the results of the $F_0$ analysis, a duration analysis is conducted for the Swedish data. The results of these $F_0$ and the duration analyses have also been presented in Ambrazaitis (2009b) and Ambrazaitis (2009a). Prior to the presentation of these investigations (6.3 to 6.5), a small-scale pilot study is reported in 6.2, which is based on materials from the Swedish PF-Star corpus (Beskow et al. 2004).

## 6.1.2 Contrastive topic and hat pattern

As a secondary object of study, this chapter treats the signalling of a *contrastive topic* (cf. 2.3.2), or more specifically, an information structure containing a *contrastive topic* (early in the sentence) and a *narrow focus* (late in the sentence). A typical realisation of a *contrastive topic plus focus* (CTOP+FOC) structure in German is a pre-nuclear rising accent for the CTOP, and a nuclear falling accent for the focus, which are connected by a high plateau (cf. 2.3.2). This *hat pattern* has been listed as a typical basic pattern of German (3.2.3). The present study is thus a complement to the study in Chapter 5, where only five of the six basic pattern types of German were treated.

Hansson (2000, 2001) discussed different cases of phrase-initial focal accents (H-) in spontaneous Swedish. While she could relate some of the initial focally accented (H-) constituents as *narrow foci*, an H- occurs also on topic constituents, which typically represent given information, among them *contrastive topics*. A common feature of these initial H- accents is a delayed $F_0$ peak, reached later than that typically observed for final H- accents, sometimes even in the following

word. In a perception test, Horne et al. (2001) showed that listeners in fact prefer a late-timed H- in phrase-initial contextually given words, as compared to an earlier timing or a lack of an H-. Hansson (2000, 2001) relates the late-timed initial H- to the hat pattern in German, which is generally assumed to signal a cohesion between the two involved accented words, and suggests that also the delay of the initial H- in Swedish may perhaps be interpreted as a coherence signal, which indicates that the initial (H-) accented word and the information that follows belong to the same prosodic phrase. Thus, Hansson (2001) concludes that it is unclear whether the late timing is a specific characteristic of a CTOP, or a more general feature of phrase-initial focal accents. One goal of this sub-study is to revisit this issue.

A second goal is to test whether the focus in a CTOP+FOC structure is signalled differently from a focus that is not preceded by a contrastive topic. As discussed in 3.2.3, it can be expected that German speakers would use a hat pattern for signalling a CTOP+FOC structure, in which an *early fall*, similarly as in a confirmative focus, is used to signal the focus, despite the fact that in a CTOP+FOC, the focus typically presents new information. The usage of the *early fall* for new-information signalling in a hat pattern may have a perception-based explanation, as suggested by Kohler (1991a) and explained in 3.2.3. If also Swedish speakers would use an *early fall* in a CTOP+FOC structure, this would, obviously, imply a strong evidence for the L- hypothesis.

## 6.2    Pilot study: the Swedish PF-Star corpus

### 6.2.1    Corpus

In the context of a project on multisensory interaction research (PF-Star), Beskow et al. (2004) collected a Swedish corpus which is appropriate for a preliminary test of the L- hypothesis. This corpus (henceforth, the PF-Star corpus) has been used in several studies on facial movements (e.g. Beskow et al. 2006), but so far not in studies on phonological issues of intonation.

A subset of the PF-Star corpus (corpus 2 in Beskow et al. 2004) is very similar to our Corpus B, since it contains three-word sentences, which were elicited with varying narrow focus on all three words (Beskow et al. 2006). These sentences were spoken in different (acted) "expressive modes", including *certain*, *confirming*, *questioning*, *uncertain*, *happy*, *angry* and a *neutral* version. Relevant for the present study are the modes *confirming* and *neutral*, which correspond to our conditions CONFIRM and NEWINFO.

The most salient differences between the PF-Star corpus and the (Swedish part of) Corpus B that are relevant for the present study may be summarised as

Table 6.1: The distribution of accent I and accent II words in the 15 test sentences from the PF-Star corpus.

| Word accent | Position | | | total |
|---|---|---|---|---|
| | initial | medial | final | |
| accent I | 11 | 2 | 12 | 25 |
| accent II | 4 | 13 | 3 | 20 |

follows.[1] First, the PF-Star corpus contains only one speaker, as opposed to nine speakers in Corpus B. Second, the material in the two corpora is structured differently, in several respects: (a) In the relevant subset of the PF-Star corpus containing the confirmative mode, there are 15 different test sentences, but no repetitions; in Corpus B, there is only one test sentence, but five repetitions in each condition. (b) The test sentence used in Corpus B consists of a noun phrase (NP), a verb, and a prepositional phrase (PP). This grammatical structure is not present among the test sentences of the PF-Star corpus, where sentences containing a subject NP, a verb, and a direct object NP are predominating (11 of 15 sentences). (c) The 15 test sentences of the PF-Star corpus are phonetically more varied than the test sentence of Corpus B, which consists almost exclusively of voiced speech sounds (cf. 4.2). (d) Corpus B is restricted to accent I, while in the PF-Star corpus, both accent I and accent II are represented, although not in a systematic manner. Table 6.1 displays the distribution of the word accents across the 45 test words (15 test sentences containing three words each) from the PF-Star corpus. The table shows that the corpus contains 25 accent I and 20 accent II words, and that accent I words are predominantly used in the initial and final positions, while in the medial position (i.e. the verb), an accent II word is used in all but two sentences. The table does not reveal exactly how the word accent types are combined in the test sentences. The most frequent combination is I–II–I (i.e. initial and final word: accent I, medial word: accent II), which occurs in 7 of the 15 sentences. Other combinations are II–II–I (4 times), I–II-II (2), I–I–II (1), and I–I–I (1); the latter corresponds to the structure of the test sentence of Corpus B.

Finally, a crucial difference between the corpora is the method of elicitation. As described in 4.2, for Corpus B, both the focus *location* and the focus *usage* (confirmative vs. plain/ new-information) were elicited by means of context ques-

---

[1]There are further general differences which are irrelevant for the present study: First, the PF-Star corpus is a multimodal corpus comprising audio and video (facial movement) data and second, as already mentioned, the PF-Star corpus contains further "expressive modes" beyond the two contained in Corpus B.

tions. In contrast, in the PF-Star corpus, no context questions were used (Beskow, personal communication). Instead, for the elicitation of narrow focus, one word in each orthographically presented test sentence was underlined and the speaker was asked to put the main stress on the underlined word. In addition, the speaker was explicitly instructed to render the test sentence in the different expressive modes.

## 6.2.2   Data analysis

The subset of the PF-Star corpus considered in this pilot study thus consists of 90 recordings (15 sentences $\times$ 3 narrow focus locations $\times$ 2 modes) by one speaker. In this corpus of 90 utterances, the variation of patterns associated with the word in narrow focus was investigated.

The analysis undertaken here is restricted to an auditory and visual inspection of $F_0$ contours by the author using Praat (Boersma and Weenink URL). Each utterance was categorised according to the $F_0$ pattern of the word in narrow focus, which was classified as either being associated with a *high+fall* (H-), or with an *early fall* (i.e. the hypothesised L-).

The PF-Star corpus also contains accent II words, but so far, no data on accent II words in a confirmation have been considered. However, it has been discussed in 5.4.2 that the *early fall* in connection with accent I has the same general pattern shape as a non-focal accent I. Hence, an *ad hoc* hypothesis for accent II is that a focussed accent II in a confirmation (H*+L L- according to the L- hypothesis) will have a tonal pattern similar to a non-focal accent II (H*+L). That is, in accent II, the $F_0$ peak in the *'early fall'*[2] will have a later timing than in accent I, as typical for non-focal word accents.[3]

Hence, a word was classified as being accented with an L- if its stressed syllable is associated with a predominantly falling (accent II) or low (accent I) $F_0$ pattern, and without any further crucial $F_0$ rise that would reflect an H- in the stressed (in the case of accent I) or post-stress (accent II) syllable.

## 6.2.3   Results and discussion

In the *neutral* context, the word in narrow focus was found to be produced with a *high+fall* pattern (H-) in 100% of the cases, as expected. In the confirming con-

---

[2]The pattern label *early fall* is inspired by the German H+L* pattern, where the pitch peak is located *early*, i.e. in the pre-stress syllable. The label seemed to be adequate also for Swedish accent I (H+L*), but, of course, it might be less appropriate in connection with accent II. However, the label is to be regarded as a rough phonetic description for the $F_0$ pattern resulting from a word accent (either I or II) combined with the hypothesised L-. Compared to an H- in accent I, the fall in an L- is *early* even in connection with accent II.

[3]However, this claim is to be modified on the basis of the results of chapter 7.

Table 6.2: Frequency of occurrence of L- accents *early fall* in connection with narrow focus in the *confirmative* utterances of the PF-Star corpus, also shown for each focus position and word accent category separately.

| Narrow focus position | Accent I | Accent II | Total | % |
|:---:|:---:|:---:|:---:|:---:|
| initial | 2/11 | 1/4 | 3/15 | *20.0* |
| medial | 0/2 | 4/13 | 4/15 | *26.7* |
| final | 9/12 | 1/3 | 10/15 | *66.7* |
| Total | 11/25 | 6/20 | 17/45 | *37.8* |
| % | *44.0* | *30.0* | *37.8* | |

text, the *high+fall* (H-) was the dominant pattern, but the *early fall* also occurred in 17 of the 45 cases (37.8%). Table 6.2 displays the occurrence of the *early fall* in more detail, separately for word accent categories and position in the utterance. The table shows that an *early fall* (L-) was found both in accent I (44%) and accent II words (30%). Furthermore, the *early fall* was most often produced in connection with final focus, where it in fact was more common than the *high+fall* (66.7% *early fall*). To compare, the *early fall* was used by 5 of the 9 Swedish speakers in the data of chapter 5 (in 28.9% of the cases for all speakers) in the condition CONFIRM (cf. Table 5.5).

The difference between the two pattern types (*early fall* vs. *high+fall*) is generally easy to recognise in the data as exemplified in Figure 6.1 for initial focus (accent I). In 6.1a (*neutral*), $F_0$ in the initial, focussed word is low in the pre-stress and rising in the stressed syllable, while in Figure 6.1b (*confirmative*), $F_0$ is high in the pre-stress and falling in the stressed syllable.

In both panels (a) and (b) of Figure 6.1, the post-focal accent II on the verb *körde* 'drove' shows its typical realisation: a fall early in the stressed vowel (H*+L), but no focal accent (H-) which would be expected in the post-stress syllable in a focal context. The post-focal accent I pattern of the final accent I word *försiktigt* 'carefully' – a fall onto the stressed vowel – is visible clearly in (a), but less clearly in (b); $F_0$ on the final words was generally difficult to extract for this speaker due to creaky voice.

Hence, the utterance in Figure 6.1b illustrates a case relevant for testing the L- hypothesis, namely an utterance, which consists of several accented words, but which does not contain a single H- accent. The question thus is whether such an utterance should be modelled as consisting of three non-focal accents, i.e. as

- (H+L*) (H*+L) (H+L*) L%
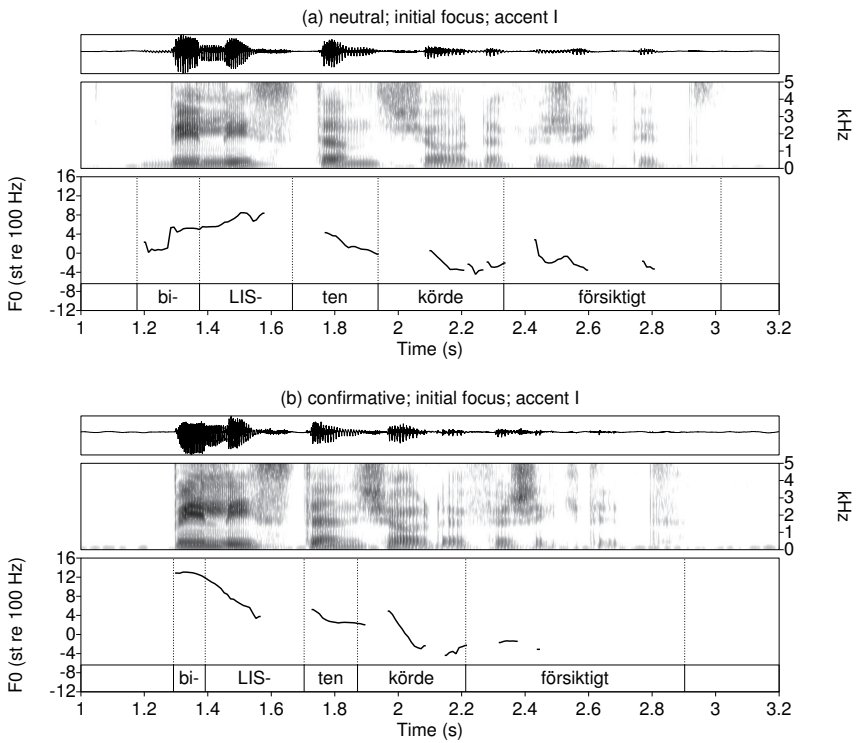
Figure 6.1: Initial focus in accent I: produced (a) with a *high+fall* in a *neutral mode* and (b) with an *early fall* in a *confirmative* mode. Test sentence: *Bilisten körde försiktigt*, broad phonetic transcription: [biˈlistən ˈɕœːɖə fœˈʂiktit] 'The driver drove carefully'. The figures show wave forms (top), spectrograms (mid), $F_0$ tracks, and an orthographic transcription, with capitals indicating the nuclear accent (bottom).

(where prosodic words are indicated by parentheses) or whether one of the accents (the first in this case) nevertheless is more prominent than the other two, rendering the accented word *narrowly focussed*, and hence prominent at the utterance level. In this case, the L- hypothesis would be strongly supported, and the more adequate representation of the sentence would be

- (H+L\* **L-**) (H\*+L) (H+L\*) L%

Although no perception experiment has been performed with the present data, it can be informally noted that the speaker of the PF-Star corpus has usually signalled narrow focus very clearly, both in the neutral and the confirmative data, and in the latter case, both in connection with a *high+fall* and with an *early fall*. This perceptual impression is typically also manifested in the acoustic data. For instance, in Figure 6.1b, it can be seen that the $F_0$ fall in the focussed word has a wider range than the corresponding $F_0$ falls in the two post-focal words.

The difference between a focal and a non-focal falling $F_0$ pattern is illustrated for all three positions in a sentence in Figure 6.2. In the panels (a), (b), and (c), respectively, the most prominent $F_0$ fall is observable in the initial (a), medial (b), and (c) final word in narrow focus. The figure also indicates that the focal and non-focal words also differ in duration. In the present example, the medial and the final word are longer when focal than when non-focal. This tendency is seen also for the initial word when only the stressed syllable is taken into account.

This brief review of the confirmative utterances taken from the Swedish PF-Star corpus has provided an initial, strong support for the L- hypothesis. To summarise, the *early fall* pattern was used by the speaker in confirmations as an alternative to the (default) *high+fall* pattern. A comparable usage of the *early fall* has also been observed in chapter 5. However, the present pilot study has extended the finding of an *early fall* in confirmations to longer sentences containing more than a single accented word. Moreover, the speaker of the PF-Star corpus used the *early fall* to signal *narrow focus* in a confirmation. Hence, besides a *high* sentence accent (H-), it is plausible to assume a *low* sentence accent (L-) for Swedish. Moreover, both could be regarded as *focal* accents, used in different pragmatic contexts: focus on *new* vs. focus on *given* information.

However, there are some possible caveats due to the structure of the corpus as described above. First, the results so far are based on only one speaker. Second, it might be discussed whether the method of elicitation used for the PF-Star corpus (cf. above) provides a sufficiently high degree of ecological validity.

The study presented in the following sections represents an attempt to replicate the results of this pilot study. However, the study treats both German and Swedish and is restricted to accent I. Moreover, a goal in designing the material was to increase the validity of the data as compared to the PF-Star corpus, both by

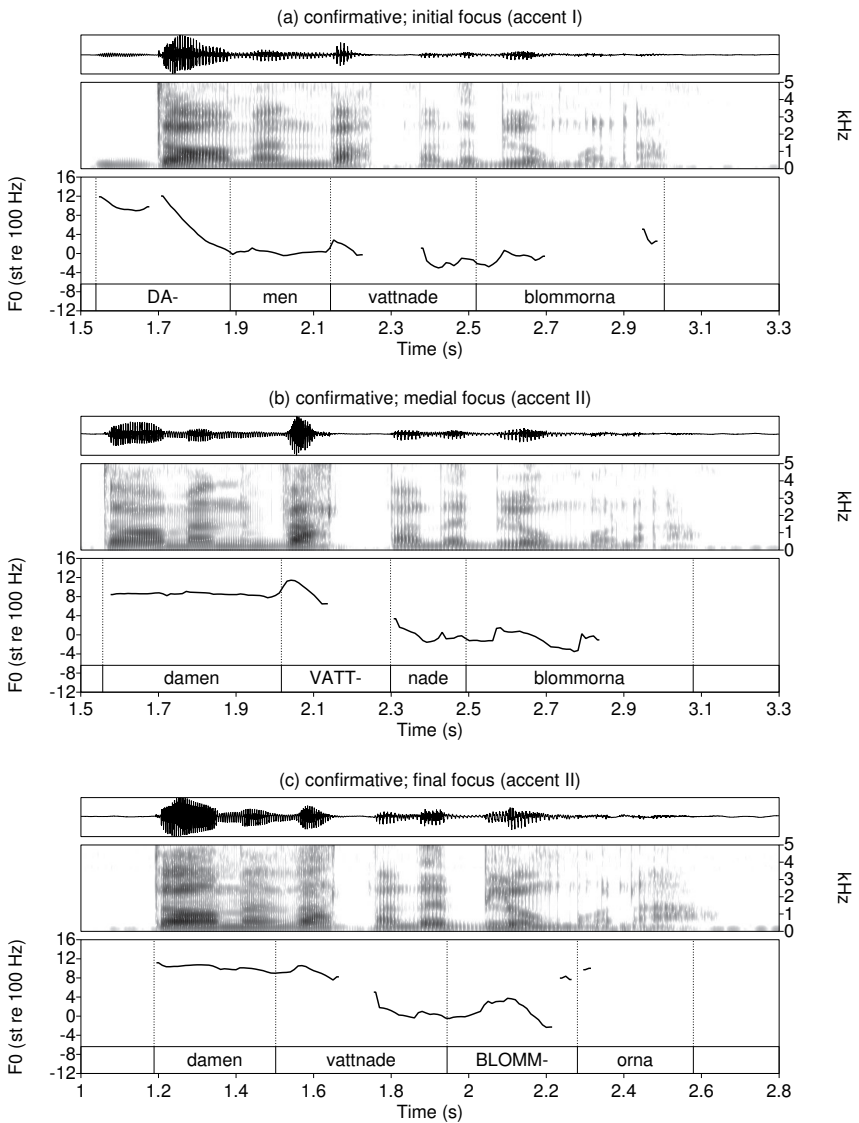Figure 6.2: (a) Initial, (b) medial, and (c) final focus produced with an *early fall* in a *confirmative mode*. Test sentence: *Damen vattnade blommorna*, broad phonetic transcription: [ˈdɑːmən ˈvatːnadə ˈblumːɔɳa] 'The lady watered the flowers'. The figures show wave forms (top), spectrograms (mid), $F_0$ tracks, and an orthographic transcription, with capitals indicating the nuclear accent (bottom).

increasing the number of speakers (per language) from one to nine, and by apply-
ing a more conversation-like elicitation method. A drawback, admittedly, is the
usage of only one test sentence, which was conditioned by the comparative aspect
of the study (cf. 4.2).

## 6.3 Method

This study is based on Corpus B. The test sentence used in this corpus is *Wallander
förlänger till november* for Swedish and *Wallander verlängert bis November* for
German ('Wallander is continuing until November'). In the case of a confirmation,
the test sentence was preceded by *ja* ('yes'). Dialogue contexts were constructed
in order to elicit the test sentence with narrow focus on the first, second, or third
content word, in each case both as a new-information response (NEWINFO), and as
a confirmative response (CONFIRM), resulting in six test conditions. An additional
seventh condition is a CTOP+FOC structure. The corpus of this study thus contains
630 utterances in total (7 conditions × 5 repetitions × 18 speakers (9 per langauge;
see Chapter 4 for a detailed description of the elicitation method, including the
context questions, the recording procedure and the subjects).

## 6.4 $F_0$ patterns – German and Swedish

The goal of this sub-study is to provide an overview of the most salient similarities
and differences between the obtained $F_0$ patterns.

### 6.4.1 Data analysis

The $F_0$ analysis consisted of an auditory and visual inspection of $F_0$ contours by
the author using Praat and a categorisation primarily according to the tonal pattern
(*high+fall* vs. *early fall*) associated with the word that was judged to be focally
accented. In the following text, tables and figures, the *high+fall* and the *early
fall* are also referred to by the simple tonal labels **H** and **L**, which are used as
an abbreviation for the tonal representation of the two pattern types introduced for
Swedish on the one hand (**H-**, **L-**), and for German on the other hand (**H***, **H+L***).

For the purpose of visual comparison and data presentation, $F_0$ contours were
normalised according to the method described in 4.2.3. For that, the utterances
were segmented into five segments, here given in a broad phonetic transcription
illustrating a possible Swedish realisation, where parentheses indicate elements
that are often elided: 1. [ʋaˈlandə(ɪ)], 2. [f], 3. [(œ)ˈlɛŋːə(ɪ)], 4. [tʰɪ(l)] and 5.

[nʊˈvɛmbəɹ]. Segments 1., 3., and 5. represent the three content words, and segments 2 and 4 were excluded from the analysis. The boundaries were set at the onsets and offsets of the initial and final sonorant portion of each content word. Each of the three content words was represented by 10 temporally equidistant $F_0$ measurements. Word-based, normalised mean contours, based on several repetitions by several speakers, were calculated in order to illustrate the general characteristics of the dominant patterns found for the test conditions. In these plots, being word-based, information on syllable boundaries is absent, but the inspection and classification of the data was carried out with reference to the stressed syllables of the test words. That is, in the categorisation of words as being accented with an *early fall* or a *high+fall*, the pitch relations of the pre-stress, the stressed, and the post-stress syllables were decisive, as in Chapter 5 (cf. Table 5.3).

## 6.4.2 Results and discussion

This section presents and discusses the patterns produced by the German and Swedish speakers in the seven conditions, primarily based on the word-based, time-normalised mean $F_0$ contours. Note that these contours are abstractions from the actual $F_0$ patterns. In particular, pauses and all duration information are eliminated. Furthermore, no attempt to perform a phrasal analysis of the recorded utterances was made. Informally, it can be noted that in most cases the test sentence has been produced as a single intonation phrase.

Tables 6.3 and 6.4 display absolute and relative frequencies of pattern types produced by the German (Table 6.3) and the Swedish (Table 6.4) speakers in the seven conditions. In some cases, it may be speculated that a produced pattern is inappropriate for a specific condition. It is difficult to judge on this issue without having performed perception experiments. However, it is likely that some mispronunciations have occurred, since the elicitation task was rather demanding. The presentation of the results in the following paragraphs focuses on the most dominant pattern types, which are displayed in Figures 6.3, 6.4, and 6.6 as time-normalised mean $F_0$ contours.

**NEWINFO and focus location in German and Swedish**

The patterns produced for the different elicited focus positions in the NEWINFO condition are in line with the expectations. Both German and Swedish speakers have marked the intended focused word with a *high+fall* (cf. H1, H2, H3, the numbers referring to the focused word, in Tables 6.3, 6.4; Figures 6.3, 6.4). According to the Lund model for Swedish, this pattern reflects a focally-accented accent I word ((H+)L* H-). For German, the pattern has been represented as an H* accent in GToBI.

Table 6.3: Relative frequencies (%) of $F_0$ patterns produced by **German** speakers. $F_0$ labels: H = *high+fall*, L = *early fall*, numbers refer to the position in the test sentence; e.g. L2 = *early fall* produced on the medial word (*verlängert*). Superscript: absolute frequency; number of speakers who produced the pattern in at least one of five repetitions. Example: **88.9**$^{40;9}$ = pattern produced at least once by all 9 speakers, in 40 of 45 cases (88.9%) in total. Relative frequencies > 15% are printed in bold face.

| Condition | $F_0$ pattern type | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | *H1* | *H2* | *H3* | *L1* | *L2* | *L3* | *other* |
| NEWINFO-1 | **100**$^{45;9}$ | | | | | | |
| NEWINFO-2 | | **93.3**$^{42;9}$ | | | 6.7$^{3;2}$ | | |
| NEWINFO-3 | | | **42.2**$^{19;6}$ | | | **55.5**$^{25;7}$ | 2.2$^{1;1}$ |
| CONFIRM-1 | **22.2**$^{10;3}$ | | | **26.7**$^{12;3}$ | 4.4$^{2;2}$ | **44.4**$^{20;5}$ | 2.2$^{1;1}$ |
| CONFIRM-2 | | **15.6**$^{7;4}$ | 2.2$^{1;1}$ | | **53.3**$^{24;8}$ | **28.9**$^{13;5}$ | |
| CONFIRM-3 | | 2.2$^{1;1}$ | 6.7$^{3;2}$ | | 2.2$^{1;1}$ | **88.9**$^{40;9}$ | |
| CTOP+FOC | | 2.2$^{1;1}$ | 2.2$^{1;1}$ | | 2.2$^{1;1}$ | **91.1**$^{41;9}$ | 2.2$^{1;1}$ |

Table 6.4: Relative frequencies (%) of $F_0$ patterns produced by **Swedish** speakers. $F_0$ labels: H = *high+fall*, L = *early fall*, numbers refer to the position in the test sentence; e.g. H1 = *high+fall* produced on the initial word (*Wallander*). Superscript: absolute frequency; number of speakers who produced the pattern in at least one of five repetitions. Example: **88.9**$^{40;9}$ = pattern produced at least once by all 9 speakers, in 40 of 45 cases (88.9%) in total. Relative frequencies > 15% are printed in bold face.

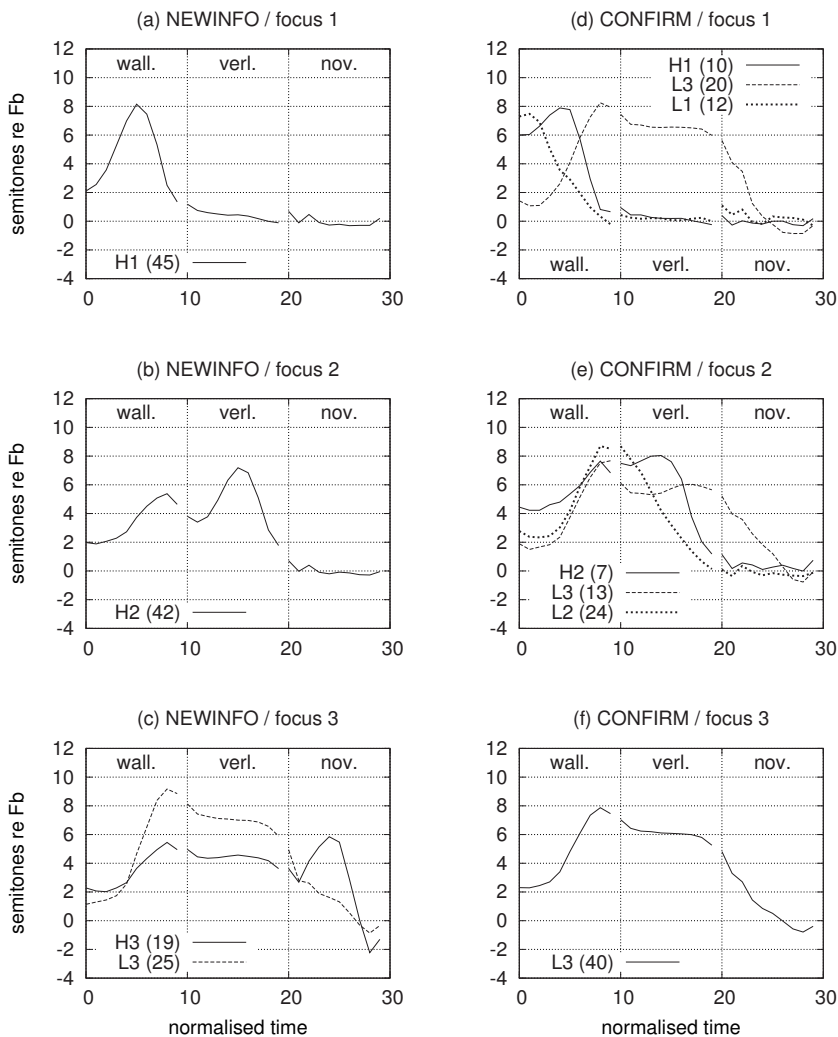| Condition | $F_0$ pattern type | | | | |
| --- | --- | --- | --- | --- | --- |
| | *H1* | *H2* | *H3* | *L2/3* | *other* |
| NEWINFO-1 | **88.9**$^{40;9}$ | | 11.1$^{5;4}$ | | |
| NEWINFO-2 | | **88.9**$^{40;9}$ | 4.4$^{2;2}$ | | 6.7$^{3;1}$ |
| NEWINFO-3 | | | **97.8**$^{44;9}$ | 2.2$^{1;1}$ | |
| CONFIRM-1 | **84.4**$^{38;9}$ | 4.4$^{2;2}$ | 4.4$^{2;2}$ | | 6.7$^{3;1}$ |
| CONFIRM-2 | 2.2$^{1;1}$ | **44.4**$^{20;7}$ | **15.6**$^{7;5}$ | **37.8**$^{17;8}$ | |
| CONFIRM-3 | 2.2$^{1;1}$ | 2.2$^{1;1}$ | **60.0**$^{27;9}$ | **35.6**$^{16;6}$ | |
| CTOP+FOC | 6.7$^{3;1}$ | | **91.1**$^{41;9}$ | 2.2$^{1;1}$ | |

Figure 6.3: Mean $F_0$ contours of the most commonly produced patterns (the cases printed in bold face in Tab. 6.3; $n$ in parentheses) by **German** speakers. Time (x-axis) is normalised; the scale indicates the number of measurements; breaks in the curves and vertical lines symbolise word boundaries. Semitones (y-axis) relate to an approximation of the speakers' base $F_0$.

Figure 6.4: Mean $F_0$ contours of the most commonly produced patterns (the cases printed in bold face in Tab. 6.4; $n$ in parentheses) by **Swedish** speakers. Time (x-axis) is normalised; the scale indicates the number of measurements; breaks in the curves and vertical lines symbolise word boundaries. Semitones (y-axis) relate to an approximation of the speakers' base $F_0$.
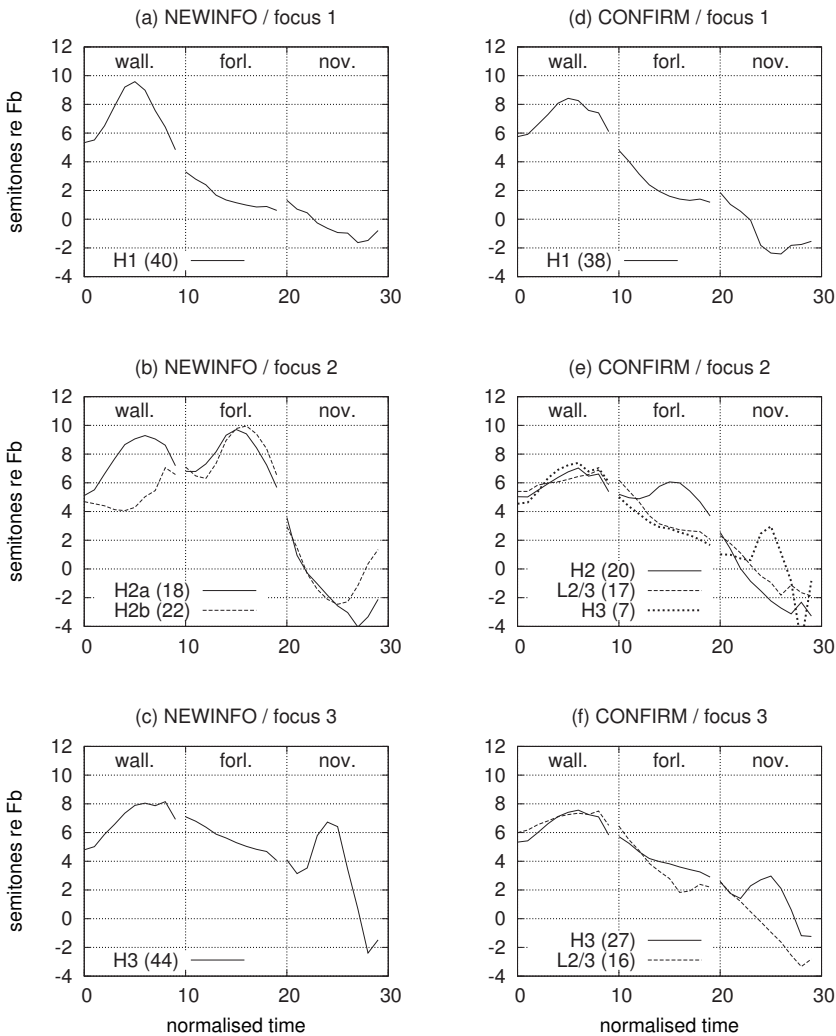
However, when narrow focus was elicited on the final word, the *high+fall* was only one of two options for German speakers, the alternative being an *early fall* (L3). This fall on the final word was preceded by a salient rise on the initial word and a high plateau on the medial word, i.e. a *hat pattern* ('t Hart et al. 1990; cf. 3.2.3). The H3 and the L3 seem to be alternatives that both do well in signalling a *plain* focus on the final word.

In the Swedish version of the H1, there is a falling $F_0$ movement on the medial and the final word each (Figure 6.4a), where $F_0$ is rather flat in the German productions (Figure 6.3a). However, especially regarding the final word, the Swedish curve should be interpreted with care, since it is much more influenced by noise than the German one; the Swedish speakers generally used creaky voice more often than the German speakers. Yet, a fall on a post-focal word in Swedish can be explained as the realisation of a non-focal accent I word (H+L*).

For both languages, there was an additional rise on the initial word even when focus was on the medial or the final word. For Swedish, phrase-initial, pre-focal rises have been analysed as additional H- accents (Bruce 1982a; Horne 1991; Horne et al. 2001; Hansson 2001; Myrberg 2009), with various suggestions concerning their function. Alternatively, these rises have been interpreted as left-edge boundary markers of main clauses (Roll 2006; Roll et al. 2009). Roll (2006) found that this boundary tone is realised as a high tone on the last syllable of the initial word. However, two variants of an initial rise occurred in the Swedish data of this study, drawn separately for medial focus in Figure 6.4b, indicating that there might be two different types of initial rises in Swedish, one representing an initial H- accent, the other representing an initial boundary tone. We will return to this issue in connection with the discussion of contrastive topics below.

In summary, *plain* focus (i.e. focus in the NEWINFO condition), is signalled intonationally in a similar way in Swedish (accent I words) and German. A salient difference, however, is that in German, utterances with focus on the final constituent can be produced with a hat pattern, including an *early fall* in the final position.

**CONFIRM and focus location in German**

Turning to the German patterns produced in confirmations, it can be recognised that the L3 pattern, realised in NEWINFO with focus elicited on the final word, was also commonly produced in CONFIRM in all intended focus positions (Table 6.3). This case is illustrated in (1) by means of the test dialogue used for the elicitation of initial narrow focus (cf. 4.2 for the dialogues used in the remaining conditions). In (1), the nuclear accent is indicated by capitals and the pre-nuclear accent by italics; the pitch accents in parentheses represent the hat pattern (L3) found in the German data.

(1)     A:    Aber wer verlängert denn bis November? Das war doch Wallander,
              oder?
              'But who is continuing until November? It's Wallander, isn't it?'
        B:    Ja, *Wallander*(L*+H) verlängert bis NOVEMBER(H+L*).
              'Yes, Wallander is continuing until November.'

The accent pattern indicated in (1) could be interpreted either as a broad focus,
or as a narrow focus on the final word, but not as a narrow focus on the initial
word. This pattern is hence expected for final focus, but not for (intended) focus
on the initial or the medial word. It would seem that speakers who have used the
hat pattern (i.e. the L3) as a kind of *default* confirmation pattern, regardless of
intended focus location, in fact prioritised the confirmation of the whole sentence,
by means of a broad focus, over the confirmation of a specific constituent (the
initial *Wallander* in the example). Although a nuclear accent on the intended
narrow focus constituent was more expected – and in fact found in the majority
of the cases, cf. Table 6.3[4] – speaker B's reply to A's question in (1) does indeed
sound adequate, too.[5] It would thus seem that the marking of a narrow focus in a
confirmation, where all information is given, is possible (cf. below), but optional.

     However, when narrow focus was intended on the initial or the medial word,
two other patterns were quite dominant, as well. One was a *high+fall* (H1 or H2
for intended focus on the initial or medial word, respectively), as also found in
NEWINFO. The only crucial difference, for example between H1 in CONFIRM and
H1 in NEWINFO, seems to be that the onset $F_0$ level is much higher in the confir-
mations (Figure 6.3a,d). A likely explanation is that, in the case of a confirmation,
the utterance started with the word *ja* (not included in the figures), which was
often produced with a rise, thus providing a higher onset for the following word
*Wallander*.

     The similarity of the H1 patterns in NEWINFO and CONFIRM may indicate that
the speakers using this pattern prioritised focus signalling over the (prosodic) sig-
nalling of a confirmation. Since all confirmations were prefixed with the word *ja*,
one could argue that there was no actual need for an additional prosodic signalling
of confirmation. This would imply that the results for confirmations could have
exhibited less variation if the initial *ja* had been omitted. However, cases where
the intonation pattern in confirmations and assertions seemed to be similar were
rather infrequent (H1 and H2 for intended focus 1 and 2 in CONFIRM, Table 6.3).

---

[4]This is first and foremost true for the medial position (cf. the sum of the frequencies for H2
and L2 in CONFIRM-2 in Table 6.3), and the final position (L3 in CONFIRM-3). In the initial
position, a nuclear accent on the first word did not occur in the majority of the cases, but still more
often than a nuclear accent on the final word (cf. the sum of L1 and H1 with the frequency for L3
in CONFIRM-1).

[5]This qualification is based on the informal judgements of three native speakers of German.

The third pattern produced by some German speakers for focus in first or second position was an *early fall* (L1 or L2, respectively), where the $F_0$ peak occurs crucially earlier than in a *high+fall* (Figure 6.3d,e). As mentioned above for L3, the usage of an *early fall* on the focused word in a confirmation is in line with the expectations. On the whole, the results indicate that, in German, a confirmation can be encoded with narrow focus (on any of three constituents), but speakers may prioritise the signalling of *confirmation* over marking of a *narrow focus*, or vice versa.

### CONFIRM and focus location in Swedish

As regards confirmations produced by Swedish speakers, the results allow for similar conclusions as for German. One strategy to signal focus on a specific word in a confirmation was to produce a *high+fall* accent on that word, although with a lower peak than in the assertions (e.g. Figure 6.4c,f). It could be argued that the underlying tonal pattern is the same as in the assertion, namely a word accent I (H+L*) plus focal accent (H-), the latter, however, being *downstepped* in a confirmation. A lowering of the peak in a *high+fall* used in CONFIRM was also found in the Swedish data in Chapter 5.

A few cases (15.6%, cf. Table 6.4) of intended narrow focus on the medial word in CONFIRM, were classified as H3, since a *high+fall* occurred on the final word, instead of on the medial word (cf. Table 6.4 and Figure 6.4), as would have been expected for medial focus. These cases might reflect the same phenomenon as observed above for the German, namely the optional marking of a broad focus in a confirmation, although the context question elicits a narrow focus on a non-final word.

However, when focus was intended on the medial or the final word, an alternative confirmation pattern occurred quite frequently, which was labelled L2/3 (Figure 6.4e,f). This pattern has a rise on the initial word, and two falling movements, one each in the medial and the final word. This L2/3 differs considerably from the H2 or the H3 pattern. Rather than a *downstepped* H-, L2/3 appears to represent a *lack* of H-. This falling pitch pattern in confirmations is expected on the basis of the L- hypothesis derived from Chapter 5.

The L- hypothesis would be best supported if, in the present study, the falling accents that are associated with the word which was intended to be focused in confirmations would differ clearly from non-focal, i.e. pre- or post-focal, accent I realisations, as observed in the pilot study (6.2). However, the mean time-normalised contours of L2/3 in focus condition CONFIRM-2 and CONFIRM-3 are nearly identical, cf. Figure 6.4e,f, or Figure 6.5a, where some critical patterns are summarised in a single plot.

One possible interpretation of these results is that medial or final narrow focus,
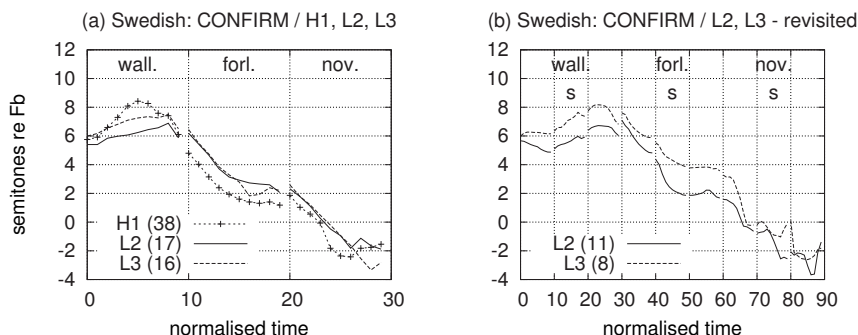
Figure 6.5: (a) Three critical patterns: H1, L2, and L3 produced by **Swedish** speakers in confirmations with intended focus on the initial (1), medial (2), or final (3) word. (b) A subset of the data for L2 and L3, containing only perceptually clear cases of L2 and L3, cf. text. s = stressed syllables. Mean $F_0$ contours ($n$ in parentheses); time (x-axis) is normalised; the scale indicates the number of measurements; breaks in the curves and vertical lines symbolise (a) word or (b) syllable boundaries. Semitones (y-axis) relate to an approximation of the speakers' base $F_0$.

in fact, were not marked at all in these confirmations, i.e. that the entire utterance would be perceived as lacking any narrow focus. Since the L2/3 pattern is even similar to the H1 (cf. Figure 6.5a), another possibility is that all patterns displayed in Figure 6.5a would be perceived with a focal accent on the *initial* word. However, informal listening (by the author) indicates that in a majority of the cases an utterance-level prominence, indicating a narrow focus, can be perceived on the medial or the final word, while the *initial* word in an L2/3 pattern was hardly ever perceived as focused.

In order to separate clear cases of narrow focus marking among the L2/3 productions from less clearer cases, two native speakers of Swedish were asked to listen to the 33 critical cases of the L2/3 pattern (17 L2, 16 L3 in Figure 6.5a) and to mark for each utterance whether it sounds more like a medial or a final focus. Medial (or final) focus was defined as 'adequate as an answer to a context question which asks for the medial (or the final) word'. The two volunteers were not told in which conditions (medial or final focus) the utterances had originally been elicited. An agreement upon the focus location between the two listeners on the one hand and the original focus condition in the data elicitation on the other hand was observed in 19 of the 33 cases. Figure 6.5b displays the curves for L2 and L3 based on these 19 cases. Unlike in Figure 6.5a, the curves in Figure 6.5b are syllable-based, rather than word-based (the syllable segmentation used is de-

scribed in connection with the duration analysis in 6.5 below). It can be seen that the curves for L2 and L3 differ more clearly in Figure 6.5b than in Figure 6.5a. In particular, the $F_0$ fall from the pre-stress to the stressed syllable of the medial (of final word) is slightly extended when the medial (or final) word is focus, as compared to the non-focal condition (i.e. when the other word is in focus).

However, the distinction between the observed $F_0$ patterns of the L2 and the L3 is still rather fine. Furthermore, the curves in Figure 6.5b should, of course, be interpreted with some care, since the amount of available data concerning the critical distinction between the L2 and the L3 is small, and only two speakers have been asked to evaluate the utterances. Heldner (2001) discussed and investigated alternative acoustic correlates of focal accents, such as duration and spectral emphasis. Such correlates, which are not represented in our normalised $F_0$ contours, might be especially important for the distinction between an H+L* L- and an H+L* pattern, considering the fine difference in $F_0$ patterns found in the present study. Therefore, Section 6.5 investigates durational patterns in the present data.

**Contrastive topic and focus in German and Swedish**

While both Swedish and German speakers applied up to three different strategies for the signalling of confirmations (cf. Figures 6.3 and 6.4), Tables 6.3 and 6.4 show that hardly any variation of pattern type occurred in the data in the signalling of a CTOP+FOC structure. Figure 6.6 displays the normalised mean curves for CTOP+FOC and NEWINFO-3 as a reference.

Both German and Swedish speakers marked the CTOP constituent (i.e. the initial word) with a rising accent. However, German speakers most commonly produced a *hat pattern* with an *early fall* signalling the final focus constituent (L3). This pattern was expected and can be represented as a prenuclear rising accent on the CTOP (L*+H) and a nuclear *early fall* on the focussed word (H+L* L%). Swedish speakers used a *high+fall* for the final focussed word (H3), which is also in line with earlier reports on focus in a CTOP+FOC structure (Hansson 2001).

As shown by Figure 6.4, both Swedish and German speakers produced basically the same pattern in CTOP+FOC as in the NEWINFO condition with final focus, although the hat pattern used by the Germans was only one of two options in NEWINFO-3 (and it was in fact more common in CONFIRM-3, cf. above). That is, also in the conditions NEWINFO and CONFIRM, where the initial words were not explicitly elicited as contrastive topics, the initial word was produced with a rising $F_0$ movement. Figure 6.6 in fact shows that the hat pattern produced by German speakers in a CTOP+FOC does not seem to differ crucially from the hat pattern produced in NEWINFO-3. For German, a clear difference between CTOP+FOC and NEWINFO-3 is thus only found for those speakers who have realised the final
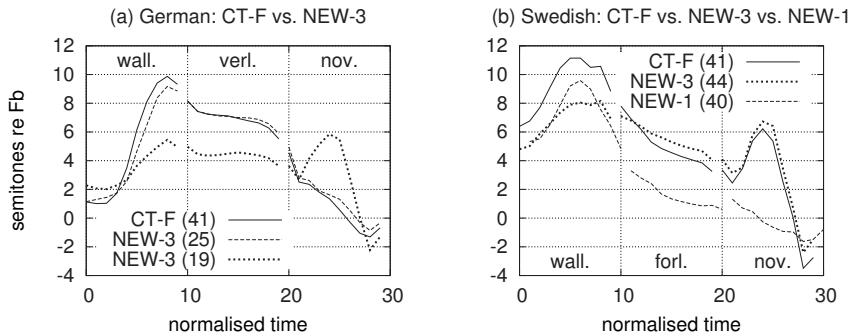
Figure 6.6: Mean $F_0$ contours ($n$ in parentheses) of the most commonly pro-
duced patterns by (a) **German** and (b) **Swedish** speakers in CTOP+FOC (CT-F)
and NEWINFO-3 (NEW-3). For Swedish (b), also NEWINFO-1 (NEW-1) is in-
cluded, cf. text. Time (x-axis) is normalised; the scale indicates the number of
measurements; breaks in the curves symbolise word boundaries. Semitones (y-
axis) relate to an approximation of the speakers' base $F_0$.

focus with a *high+fall*. Swedish speakers, however, seem to have signalled a dif-
ference between CTOP+FOC and NEWINFO-3 more clearly by means of a *raised*
peak associated with the CTOP.

It has been reported that initial focal accents (H-) may be *timed late* (Horne
et al. 2001; Hansson 2001). Moreover, Hansson (2001) discussed that this late
timing could either have the specific function of marking a CTOP, or that it could
be a feature of initial H- accents in general, possibly with the function of co-
herence signalling, similar to the high plateau in a hat pattern (cf. 6.1.2). In the
present data, in both conditions CTOP+FOC and NEWINFO-3, the initial (H-) ac-
cent seems to be timed rather *early* in the sense that the peak is reached already
in the stressed syllable. However, the peak is broad, spanning both the stressed
and the post-stress syllable, and may hence also be classified as *late* in the sense
that $F_0$ does not fall to a low level immediately after the $F_0$ peak. To compare,
the peak in the *high+fall* for initial narrow focus in NEWINFO-1 could be clas-
sified as 'non-late-timed', since a fall immediately follows the peak (cf. Figure
6.6b). However, in the present data, the 'late-timed' initial peak does not seem to
be timed later in a CTOP-FOC than in the NEWINFO-3 condition (cf. Figure 6.6b).
Thus the present results seem to support the alternative hypothesis formulated in
Hansson (2001), namely that the late timing is not a special feature of a CTOP, but
rather a more general feature of initial H- accents. Moreover, as just mentioned,
the late timing was not found in the initial narrow focus condition (NEWINFO-1).
That is, in the present data, the late timing occurs only in case the initial H- is

followed by a second, final H-. A possible interpretation of this result is that the late-timed initial H- indeed signals coherence between the initial and the final H-, similarly as the high plateau in a hat pattern in German, in line with the proposal by Hansson (2001).

Finally, it was also mentioned above that initial rises in Swedish have not only been treated as initial accents (e.g. Hansson 2001; Myrberg 2009), but also as initial boundary tones (Roll 2006; Roll et al. 2009). Since two different initial rises were observed in the NEWINFO condition in the Swedish data (cf. Figure 6.4b), it was suggested above that the two proposals concerning initial rises (initial accent vs. initial boundary tone) may be adequate at the same time, but that they in fact refer to two different phenomena. The initial accent used for the marking of a CTOP is indeed very similar to one of the rise types (H2a) shown in Figure 6.4b. While the $F_0$ peak may be timed relatively late in these rises in the sense that the $F_0$ peak is broad, a high $F_0$ level is nevertheless already reached in the stressed syllable (in the case of accent I), at least in the present data. It would hence seem appropriate to model both phenomena (i.e. the initial rise found in CTOP and one of the initial rise types found in NEWINFO: H2a in Figure 6.4b) as an initial H- accent. This initial H- might have different functions (cf. Myrberg 2009), among them the signalling of a CTOP. In the other initial rise type (H2b) found in NEWINFO (cf. Figure 6.4b), however, the stressed syllable is markedly low, and the rise starts late. This type matches the descriptions of the phenomenon that has been referred to as an initial boundary tone by Roll (2006) and Roll et al. (2009).

## 6.5   Duration patterns – Swedish

It is known for a variety of languages that prosodically focussed words (in a new-information context) are not only marked tonally, but also temporally, i.e. by means of lengthening (e.g. Bruce 1981; Heldner and Strangert 2001 for Swedish; Cooper et al. 1985 for American English; Eefting 1991 for Dutch; Cambier-Langeveld and Turk 1999 for English and Dutch; Kügler 2008 for German).[6] Moreover, Bruce (1981) suggests that increased duration is not merely an adaptation to the more complex tonal pattern, but rather a focus cue on its own, besides the tonal gesture.

The goal of this study is to examine durational patterns in the Swedish part of the present data on focus marking in NEWINFO and CONFIRM. The hypothesis is that, if narrow focus is signalled in confirmations, and if lengthening is a focus cue independent of the tonal pattern, then focal lengthening should be found not

---

[6] A more detailed review is presented in Heldner and Strangert (2001).

only in NEWINFO, but also in CONFIRM. Furthermore, it could also be the case that durational patterns differ in NEWINFO and CONFIRM, indicating that duration is also a cue to the referential status *new vs. given* of the focus constituent.

## 6.5.1 Data analysis

In order to obtain duration measurements, the utterances were segmented into 10 quasi-syllables. The boundaries between the segments were set as illustrated by the following broad phonetic transcriptions: [ʋa], [ˈland], [əɪ], [f(œ)], [ˈlɛŋː], [əɪ], [tʰɪ(l)], [nʊ], [ˈʋɛmb], [əɪ]. In the case of [ˈland] and [ˈʋɛmb], the final boundary was set at the time of the plosive burst, if present, or at the onset of the post-stress vowel.

It has been shown for Swedish that focal lengthening is non-linear, in that the stressed syllable is lengthened more than the unstressed syllables (Heldner and Strangert 2001). Therefore, durational patterns were analysed on two levels, first, taking into account entire word durations, second, concentrating on stressed syllables only. In both cases, the analyses focussed on the three content words and hence disregarded the function word *till*.

For each word, two repeated-measures ANOVAs were calculated, one with *word duration* as the dependent variable, the other for *stressed syllable duration*. In each of the six ANOVAs, there were three factors: STATUS (= referential status of the focussed information, two levels: new (NEWINFO), given (CONFIRM)), FOCUS (= narrow focus location, three levels: focus on initial, medial, final word), and finally REPETITION (five repetitions, i.e. five levels).

All data were included in these six ANOVAs, irrespective of possible mispronunciations, or the intonation patterns produced (cf. the two strategies for confirmations, Figure 6.4e,f), in order to obtain a general picture of the effects of FOCUS and STATUS on duration. However, the major issue is whether focus in confirmations may be signalled by an *early fall* pattern. Therefore, in a second step, durational patterns were looked at with respect to the classification of $F_0$ patterns made in the $F_0$ analysis (cf. Table 6.4).

## 6.5.2 Results and discussion

Figure 6.7 displays mean durations of the three test words and the corresponding stressed syllables for the six conditions (three narrow focus positions in each condition NEWINFO and CONFIRM). The figure shows that the final word (*november*) is generally produced relatively long even when unfocussed, i.e. longer than medial or initial unfocussed words, reflecting the well-known phenomenon of final lengthening. Moreover, the medial word (*förlänger*) is generally produced relatively short. Figure 6.7 also shows that each word is produced longer when

Table 6.5:  Results of the six repeated-measures ANOVAs:  degrees of freedom (Greenhouse-Geisser corrected where sphericity cannot be assumed), $F$-values, and $p$-values. Factor REPETITION was never significant; no interactions besides the one shown were significant, an exception being FOCUS*REPETITION for [nʊˈvɛmbəɹ] ($F(8, 64) = 2.21$; $p = .038$).

| | [ʋaˈlandəɹ] | [fœˈlɛŋːəɹ] | [nʊˈvɛmbəɹ] |
|---|---|---|---|
| STATUS | $F(1, 8) = 9.15$ $p = .016$ | $F(1, 8) = 17.36$ $p = .003$ | $F(1, 8) = 36.33$ $p < .001$ |
| FOCUS | $F(1.13, 9.07) = 19.25$ $p = .001$ | $F(2, 16) = 39.13$ $p < .001$ | $F(2, 16) = 34.57$ $p < .001$ |
| STATUS × FOCUS | n.s. | $F(2, 16) = 20.82$ $p < .001$ | $F(2, 16) = 28.03$ $p < .001$ |
| | [ˈland] | [ˈlɛŋː] | [ˈvɛmb] |
| STATUS | $F(1, 8) = 26.59$ $p = .001$ | $F(1, 8) = 73.23$ $p < .001$ | $F(1, 8) = 25.99$ $p = .001$ |
| FOCUS | $F(1.03, 8.26) = 16.57$ $p = .003$ | $F(2, 16) = 36.92$ $p < .001$ | $F(2, 16) = 25.94$ $p < .001$ |
| STATUS × FOCUS | $F(2, 16) = 3.75$ $p = .046$ | $F(2, 16) = 22.59$ $p < .001$ | $F(2, 16) = 31.06$ $p < .001$ |

it is in focus than when it is pre-focal or post-focal (i.e. when another word is focussed).  These general tendencies have also been reported by Heldner and Strangert (2001). Since the material investigated in Heldner and Strangert's (2001) experiment 1 has a similar structure as our test material, their results constitute a suitable reference for the present study. While Heldner and Strangert (2001) take into account both accent I and II, we are only concerned with accent I. However, an additional factor in the present study, compared to Heldner and Strangert's (2001) experiment 1, is the referential status of the focussed material. Figure 6.7 shows that the focal lengthening effect can be observed for both NEWINFO and CON-FIRM, although the effect appears to be smaller in CONFIRM than in NEWINFO. For unfocussed words, there seem to be no duration differences between the two referential status conditions.

These observations are generally supported by the inferential statistics (cf. Table 6.5), although most clearly for the medial word: A significant effect was found for the factors STATUS and FOCUS, as well as for the interaction of the two factors, both for word duration ([fœˈlɛŋːəɹ]) and stressed syllable duration ([ˈlɛŋː]); no other significant effects were found for the medial word. According

**(a) Word durations**
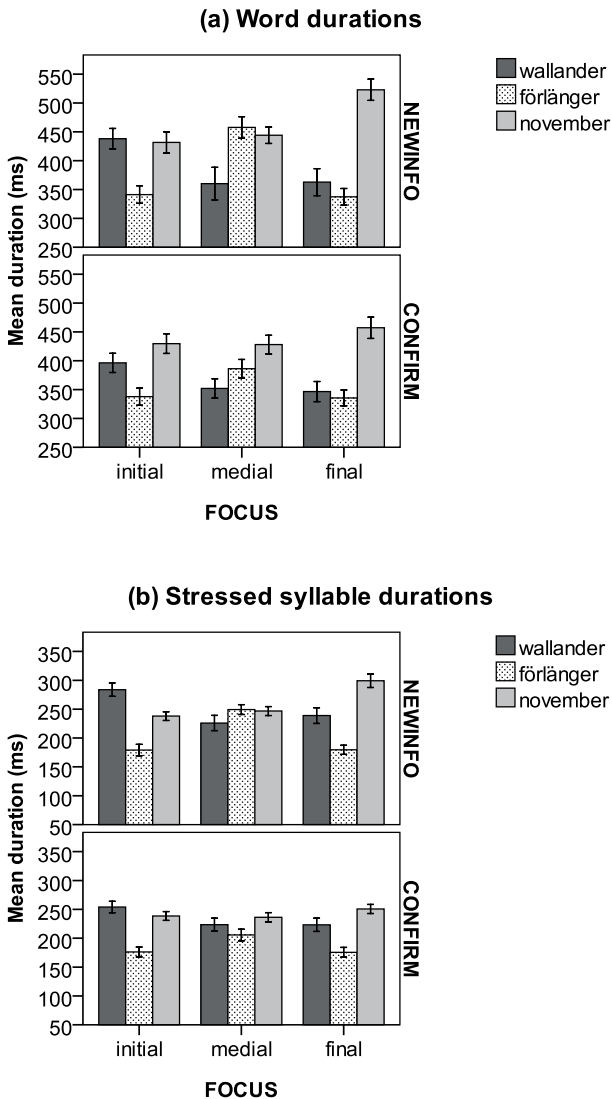


**(b) Stressed syllable durations**



Figure 6.7: Mean durations in ms of the three test words in NEWINFO and CON-FIRM with initial, medial, and final narrow focus; (a) entire words, (b) stressed syllables only; pooled across 45 repetitions by 9 speakers; error bars represent 95% confidence intervals.

to an analysis of simple main effects (with Bonferroni correction), [fœˈlɛŋːɔɪ] and ['lɛŋː] were realised with a longer duration in NEWINFO than in CONFIRM[7] only when the medial word itself was in focus. In NEWINFO, both the word and the stressed syllable were longer when the word was in focus than when another word was in focus.[8]  In CONFIRM, the general result was the same.[9]  There were no significant differences between the two non-focal conditions.

The situation is similar for the final word, the major difference in the test results being that the interaction of FOCUS and REPETITION was significant for word durations (cf. Table 6.5). Resolving this interaction shows that significant differences between repetitions only occur for final focus, and furthermore, that they seem to be restricted to confirmations. A possible explanation is that the two different strategies of focussing the final word in confirmations (*high+fall* vs. *early fall*) are reflected in this interaction. As in the case of the medial word, an analysis of simple main effects reveals that both [nʊˈvɛmbəɪ] and ['vɛmb] were realised with a longer duration in NEWINFO than in CONFIRM[10] only when the final word itself in focus. Also, the entire word is longer when in focus than in the two post-focal conditions, where either the initial word or the medial word is in focus.[11]  The picture is, however, different when only the stressed syllable is measured. In CONFIRM, no significant differences are found for ['vɛmb] in the different focus conditions, while in NEWINFO, the duration of ['vɛmb] differs in all three focus conditions.[12]

Finally, for the initial word, the interaction of FOCUS and STATUS was not significant for word duration (cf. Table 6.5). That is, [ʊaˈlandəɪ] was produced longer in NEWINFO than in CONFIRM, both when in focus and in the two pre-focal conditions (cf. also Figure 6.7). Post-hoc tests for FOCUS show that [ʊaˈlandəɪ] is realised with a longer duration when the word is in focus than when focus is on the medial (p=.011) or final word (p=.003). However, when only the stressed syllable is taken into account, the interaction of STATUS and FOCUS is significant (cf. Table 6.5). As shown by an analysis of simple main effects, the situation is, however, more complex than for the interactions found for the other words: First, ['land] is realised longer in NEWINFO than in CONFIRM not only when the

---

[7]entire word: NEWINFO > CONFIRM (p<.001); stressed syllable: NEWINFO > CONFIRM (p<.001)

[8]entire word: focal > post-focal (p<.001), focal > pre-focal (p<.001); stressed syllable: focal > post-focal (p<.001), focal > pre-focal (p<.001)

[9]entire word: focal > post-focal (p=.016), focal > pre-focal (p=.003); stressed syllable: focal > post-focal (p=.023), focal > pre-focal (p=.001)

[10]entire word: NEWINFO > CONFIRM (p<.001); stressed syllable: NEWINFO > CONFIRM (p<.001)

[11]NEWINFO: final focus > initial focus (p<.001), final focus > medial focus (p<.001); CONFIRM: final focus > initial focus (p=.018), final focus > medial focus (p=.007)

[12]final focus > medial focus (p=.001); final > initial (p<.001); medial > initial (p=0.19)

Table 6.6: Mean focal lengthening of entire words in ms ($\Delta W$) and in % (%$W$), and of stressed syllables in ms ($\Delta S$) and in % (%$S$), calculated in relation to the non-focal conditions, for both NEWINFO and CONFIRM.

|  | NEWINFO | | | | CONFIRM | | | |
|---|---|---|---|---|---|---|---|---|
|  | $\Delta W$ | %$W$ | $\Delta S$ | %$S$ | $\Delta W$ | %$W$ | $\Delta S$ | %$S$ |
| Wallander | 77 | 18 | 51 | 18 | 47 | 11 | 31 | 12 |
| förlänger | 118 | 26 | 70 | 28 | 49 | 12 | 30 | 15 |
| november | 85 | 16 | 55 | 19 | 28 | 6 | 14 | 6 |
| Grand mean | 93 | 20 | 59 | 22 | 41 | 10 | 25 | 11 |

initial word is in focus (p=.002), but also when the final word is in focus (p=.029). Second, in NEWINFO, the duration of [ˈland] differs in all three focus conditions,[13] while in CONFIRM, [ˈland] is significantly longer in focus than in the two pre-focal conditions only[14], i.e. no significant difference is found between the two pre-focal conditions.

In order to quantify the amount of focal lengthening, Heldner and Strangert (2001) used the mean duration of the two non-focal conditions pooled, for each word, as a baseline. This choice was motivated by the result that the difference between the non-focal conditions (e.g. pre-focal and post-focal in the case of the medial word) was generally small and not significant in two of the three positions in the sentence. The situation is similar in our data, where no significant differences in word durations were found between the two non-focal conditions. (Only for the stressed-syllable durations in NEWINFO, significant differences between the non-focal conditions in case of the final and the initial word were found, cf. above). Hence, we have applied the same approach as Heldner and Strangert (2001) and calculated a mean duration across the two non-focus conditions for each word (or position), but separately for NEWINFO and CONFIRM. These six non-focal reference values were calculated for both word and stressed syllable durations. Table 6.6 displays the amount of focal lengthening both for entire words and stressed syllables, calculated on the basis of the non-focal reference values, for all three words in both conditions NEWINFO and CONFIRM.

The grand means for focal lengthening found for NEWINFO are 20% for word duration, and 22% for stressed-syllable duration (cf. Table 6.6). These values are similar, but slightly lower than the ones reported in Eefting (1991) for Dutch,

---

[13]initial focus > medial focus (p=.015); initial > final (p=.036); final > medial (p=.039)
[14]initial > medial (p=.005); initial > final (p=.016)

in Cooper et al. (1985) for American English, and in Heldner and Strangert's (2001) study 1 (25% and 30%) for Swedish. However, in a follow-up study (their study 2) using different materials and different speakers, Heldner and Strangert (2001) found a grand mean of only 12% for word durations, and notice that the exact amount of focal lengtheing is largely speaker dependent. Table 6.6 also shows that the order of magnitude of focal lengthening in CONFIRM was only 10%, i.e. smaller than in NEWINFO, but not much smaller than the figure reported in Heldner and Strangert's (2001) study 2 (12%).

The amount of focal lengthening appears also to depend on the position in the sentence. In the present data, the highest values for the focal lengthening of the word (26%) or the stressed syllable (28%) are found for the medial word in NEWINFO (cf. Table 6.6). That is, the initial and final word were lengthened less than the medial word in NEWINFO. In CONFIRM, however, the initial and medial word were lengthened by approximately the same amount (in %), while the final word tended to be lengthened least. The general tendency found in the data (including both CONFIRM and NEWINFO) is thus that the final word is lengthened less than the medial or the initial word. This result is opposite to the tendency found in Heldner and Strangert's (2001) experiment 2, but in line with Cooper et al. (1985) for American English, Cambier-Langeveld (2000) for Dutch, as well as Heldner and Strangert's (2001) experiment 1, which is structurally more comparable to the present study than their experiment 2.

In the analysis so far, all recordings were included irrespective of the variation of $F_0$ patterns produced within an experimental condition. As mentioned above, confirmations were signalled using either of two strategies, as classified in 6.4 as either H (presence of a (lowered) H- accent on the target word), or L (absence of an H- accent on the target word, i.e. presence of a hypothesised L-), cf. Figure 6.4e,f. This raises the question as to whether the focal lengthening found in confirmations (cf. Figure 6.7) is present in both strategies (i.e. using the H or the L pattern). Figure 6.8 displays the results for CONFIRM in a rearranged form, where the $F_0$ pattern is taken into account.

Figure 6.8 indicates that, first, the medial word seems to be lengthened in focus even when it is produced with an L pattern (cf. *förlänger* in conditions medial L vs. final L), and second, the focal lengthening effect still tends to be stronger when the word is produced with an H (medial L vs. medial H). However, for the final word, focal lengthening seems to be present only when the word is produced with an H. Finally, the initial word seems to be lengthened not only when it is in focus itself, but also when medial or final focus is produced with an L, as compared to medial or final focus produced with an H.

Heldner and Strangert (2001) conclude that the medial position "is least affected by factors other than the focal accents, for example final lengthening". This is in line with the present results, since the lengthening was found to be most

**(a) Word durations in CONFIRM**



**(b) Stressed syllable durations in CONFIRM**



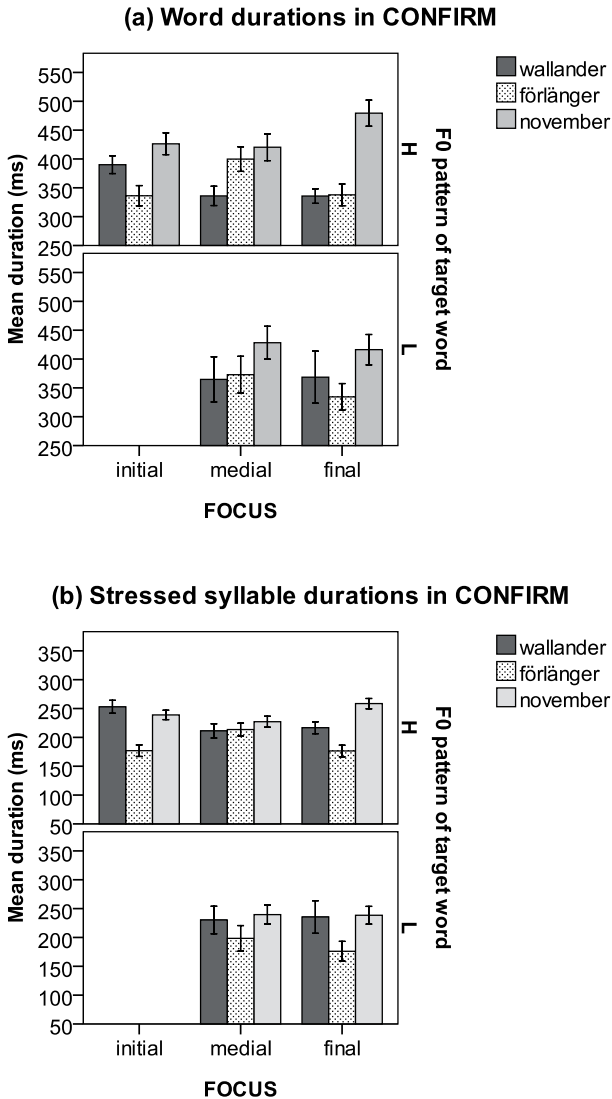Figure 6.8: Mean durations in ms of the three test words in CONFIRM with initial, medial, and final narrow focus, divided into two classes according to the $F_0$ pattern produced on the target word (H, L); (a) entire words, (b) stressed syllables only; sample sizes vary from $n = 16$ (L in final focus) to $n = 38$ (H in initial focus), cf. Table 6.4; error bars represent 95% confidence intervals.

clearly related to focus in case of the medial word. Furthermore, it seems obvious that not only the duration of the final, but also of the initial word is influenced by other factors than focus, since even if the initial word is pre-focal, its duration seems to vary depending on whether the medial or the final word is focussed (when only the stressed syllable is measured), or, in confirmations, whether medial or final focus is produced with an L or an H. More research is needed in order to reach a better understanding of these patterns. In part, durational patterns of initial words could possibly be related to the role the initial position plays in signalling phrase- or sentence prosody (Myrberg 2009; Roll et al. 2009).

## 6.6   Summary and conclusions

The primary goal of this study was to test the L- hypothesis developed in Chapter 5. For that, it was investigated whether German and Swedish speakers can signal narrow focus, by prosodic means, not only in connection with new information, but also in a confirmation. The hypothesis was that narrow focus would be signalled by means of a sentence accent in both utterance types, but that the choice of accent would be a *high+fall* (H- for Swedish) in connection with new information (condition NEWINFO), and an *early fall* (L-) in a confirmation (CONFIRM).

The analysis of $F_0$ patterns in 6.4 showed that, in NEWINFO, both German and Swedish speakers produced the expected pattern, namely a *high+fall* on the word in narrow focus. For CONFIRM, however, the situation is more complicated. In general, the parallel signalling of a confirmation and narrow focus seems fully possible, but two alternative patterns occurred in the data.

One strategy used by some (German and Swedish) speakers in confirmations was to signal focus position by means of a *high+fall*. For German, this case was rare, but when it occurred, the $F_0$ pattern did not seem to differ substantially from the corresponding pattern produced in NEWINFO. Thus, these German cases probably lack any prosodic cues for confirmation. As discussed in 6.4, this can be explained by the redundancy of prosodic cues for confirmation introduced by the test sentence used in NEWINFO, which includes an initial *ja*.

For Swedish, the case of a *high+fall* accent in a CONFIRM was quite common, but the resulting $F_0$ patterns differed nevertheless clearly from the *high+fall* produced in NEWINFO, since the $F_0$ peaks were produced substantially lower in CONFIRM than in NEWINFO. A possibility to model the lowered H- in a confirmation might be in terms of *downstep* (!H-). Of course, a perception test would be required in order to determine if this downstepped accent would suffice as a cue to a confirmation. However, the results provide at least a strong indication that the Swedish speakers using this strategy indeed have signalled focus and confirmation simultaneously, although they used other means than the one predicted by the L-

hypothesis.

The other strategy was to use an *early fall* (L) pattern as predicted by the hypothesis. First and foremost some of the German speakers have clearly used an L pattern on the word in narrow focus (either the initial, medial, or late word) in confirmations. For medial and final focus, the L was preceded by a rise on the initial word, which resulted in a hat pattern connecting the first and the medial, or final word, respectively. However, some speakers have produced a hat pattern with an L on the final word in all three intended focus positions. While this pattern is probably well-suited for signalling a confirmation, it is unlikely that it suffices to signal narrow focus on the initial or the medial word.

In the case of Swedish, a common pattern produced in confirmations was a *double fall* (above referred to as L2/3), which implies a (non-focal) rise on the initial word, and a fall on both the medial and the final word. This pattern lacks a rising focal accent (H-) and hence, it is likely that this pattern is suitable for signalling a confirmation. Furthermore, this pattern implies a fall on the focussed word, which was predicted by the hypothesis. The additional fall on the pre- or post-focal word is also expected as the non-focal realisation of accent I (H+L*).

The crucial issue concerning this pattern is whether the fall on the focused (medial or final) word would in some respect differ acoustically and perceptually from the fall on the non-focal (final or medial) word. If this were the case, then the *double fall* pattern would support the general hypothesis of a falling sentence accent in Swedish (L-). The present $F_0$ analysis using mean $F_0$ curves has provided some evidence for a distinction between the two falls (cf. Figure 6.5b), but as discussed in 6.4, the distinction is fine and the results are based on a small set of data. It can be concluded that the results of the $F_0$ analysis in 6.4 do not provide any conclusive evidence for the hypothesis that a *falling sentence accent* (L-) can be used to signal narrow focus in a confirmation, but they point in that direction.

However, additional support for the L- hypothesis was provided by the pilot study presented in 6.2, where a clear difference in $F_0$ patterns was observed between the fall produced on a narrowly focussed word in a confirmation (i.e. a hypothesised H+L* L-, for accent I) and a non-focal fall (H+L*). It was discussed in Chapter 5 that the proposed L- accent does not primarily describe an additional tonal gesture. Based on the pilot study in 6.2 and on Figure 6.5b, the realisation of the L- accent could be characterised as a 'boost' of the word-accent fall.

The results of the duration analysis provide some further support for the L-hypothesis, since they have shown that focal lengthening can be found not only in connection with new information, but also in confirmations, although the degree of focal lengthening seems to be smaller in confirmations than in new-information responses. Most importantly, the results also indicate that some focal lengthening may even be found when the target word is produced with an *early fall*. Some of the duration differences found in this study are small and probably irrelevant

from a perceptual point of view (cf. Heldner and Strangert 2001 for a discussion). However, the general tendencies found in the results suggest that an 'L- boost' is also manifested in the time domain, and hence that duration is a possible cue to focus position in confirmations. A future perception experiment should attempt to pinpoint the relevance of duration and $F_0$ cues for the signalling of an L- accent.

A remark is in order concerning the significance of the results obtained in the present chapter. Some test conditions of the present study are rather artificial. In particular, the case of a narrow focus on a sentence initial or medial constituent in a confirmation is probably rather infrequent in natural spoken communication. Consequently, the results suggest that not all speakers seemed to feel comfortable with the task of signalling a confirmation with a specific narrow focus, since they have prioritised focus signalling over the signalling of confirmation, or vice versa, which was most clearly seen in the German data. As discussed above, some German speakers have used a *hat pattern* with a nuclear accent on the final word in a confirmation, regardless of the elicited narrow focus, while others have used a nuclear H* accent on the narrow focus, however, without producing any salient differences in the $F_0$ pattern in confirmations as compared to the new-information condition. However, other speakers, both German and Swedish, have used an *early fall* pattern in order to signal narrow focus in a confirmation, as predicted by the hypothesis. It may thus be concluded that, on the one hand, the results support the hypothesis of a low (L-) sentence accent in Swedish. On the other hand, they also suggest that this accent is probably not regularly used in the specific pragmatic function simulated in this study (i.e. signalling a specific narrow focus in a confirmation).

To summarise the results of the pilot study (6.2) and the $F_0$ (6.4) and the duration (6.5) analyses of Corpus B, the data generally support the L- hypothesis, although the support could have been stronger. However, it seems to be possible to signal narrow focus on given information (i.e. in a confirmation) by means of an *early falling* sentence accent in Swedish, as in German. If we assume that, in Swedish, the observed fall is determined by the word accent in the first place,[15] then the effect of the L- may be characterised as a *boost* of the word-accent pattern. The phonetic correlates of this boost may be a tonal expansion of the fall, but also an increase in duration (focal lengthening).

It can also be noted that the L- hypothesis is generally in line with the results and conclusions of Strangert and Heldner (1994) and Heldner (1997, 1998, 2001). Strangert and Heldner (1994) observed that words which are rated high on a prominence scale by labellers of speech corpora are not always produced with a rise in $F_0$ as expected for an H- focal accent. It would be interesting to test whether some of these cases occurred in confirmations or similar contexts and

---

[15]This view, however, might need a revision with respect to the results of Chapter 7.

reflect what is here hypothesised to be an L- accent. Moreover, Heldner (1997, 1998) concludes from perception experiments that an $F_0$ rise is not a necessary cue to the location of focus.[16]

A general remark can be made on the patterns produced in confirmations in German and Swedish. If we assume that our preliminary conclusion is valid, namely that even in the case of Swedish, we can distinguish between a falling sentence accent on the medial (L2) and one on the final word (L3) in our test material, then patterns that are rather different at the surface can be characterised as underlyingly similar. For instance, the L3 is realised as a *hat pattern* in German, but as a *double fall* in Swedish. Both have a rise on the initial and a fall on the final word. The difference concerning the medial word, where the German hat pattern has a high plateau, and the Swedish pattern is falling, could be explained by the presence of word accents in Swedish, or more precisely, by the fact that, according to the Lund model, even accent I has its own tonal correlate (H+L*).

The secondary object of study has been the prosodic signalling of a CTOP+FOC. The German speakers marked the CTOP+FOC structure with a hat pattern, implying a rising accent on the CTOP, and an *early fall* on the focus. The Swedish speakers, instead, marked the CTOP and the focus with two separate H- accents. A general characteristic of the initial H- marking CTOP found in the present study is a *raised $F_0$* peak, compared to the final H- which marks the final focus, but also compared to an initial (non-focal) H- produced in the NEWINFO conditions with medial or final focus (cf. Figure 6.6 above). A second feature of the initial H- that has been reported in previous studies (e.g. Hansson 2001; Horne et al. 2001) is the *late timing* of the H- peak. As mentioned above, Hansson (2001) concluded that it is unclear whether the late timing is a general feature of an initial H-, irrespective of its function, or a more specific feature of an H- marking a CTOP. The results of the present study suggest that the late timing (in the sense of a broad, or distributed $F_0$ peak) can be found more generally in initial H- accents which do not necessarily mark a CTOP (cf. also Myrberg 2009), and that its function may be the signalling of coherence with the final, focus signalling H-.

As in the case of the *double fall* in Swedish vs. the *hat pattern* in German discussed above, the different melodic patterns found for the two languages in the marking of a CTOP+FOC structure might reflect the same underlying intonational structure. Again, the fall on the medial word between the initial and the final H-accents in a Swedish CTOP+FOC structure can be explained as a non-focal accent I pattern. Yet, the German and the Swedish realisation differ not only in the fall

---

[16]However, although Heldner's (1997, 1998) conclusion is generally in line with ours, his experiments do actually not provide a support for the L- hypothesis, since they were, of course, not designed to do so. For that, it would be necessary to test whether a *falling $F_0$* pattern (in addition with further acoustic cues such as duration) can be a sufficient cue to signal narrow focus in a confirmation.

(Swedish) vs. the high plateau (German) on the medial word, but also in the shape of the nuclear accent on the final word, which is realised as an *early fall* in the German, but a *high+fall* in the Swedish data. However, if the *early fall* in the German hat pattern can be motivated by a perceptual effect as suggested by Kohler 1991a (cf. 3.2.3), it would seem possible that even in Swedish, a *hat pattern* including an *early fall* would be used if no (word) accent intervened between the CTOP and the focus.[17] This should be tested in future research with appropriate test material.

Finally, the present chapter might also have made a contribution to the discussion of non-focal, phrase- or sentence-initial rises in Swedish (e.g. Horne et al. 2001; Hansson 2001; Roll 2006; Roll et al. 2009; Myrberg 2009), since it suggests that they might be two different types of initial rises, which have not been distinguished in the literature so far (cf. Figure 6.4b).

---

[17]Since complete de-accentuation of content words in uncommon in Swedish, as witnessed also by the present data, this hypothesis cannot be tested by the present material. However, in one exceptional case, a speaker has indeed de-accented the medial word *förlänger* and produced a *hat pattern* in the CTOP+FOC condition, including an *early fall* on the final focus.

# Chapter 7

# Nuclear pattern type and word accents in Swedish

## 7.1  Introduction

Chapter 6 focussed on the nuclear pattern contrast between a *high+fall* and an *early fall* in German and Swedish, as observed in the conditions NEWINFO and CONFIRM, and added a further pragmatic parameter, namely narrow focus location. Variations due to the Swedish word accent have, however, not been taken into account so far, with the exception of the pilot study presented in 6.2. The present chapter continues to investigate the *high+fall* vs. *early fall* contrast in NEWINFO and CONFIRM, but focusses on Swedish, and takes into account the word accent distinction in a systematic manner.

Two studies are presented, a production study (7.2) and a perception experiment, involving reaction time measurements (7.3). The production study is based on Corpus C (cf. 4.3), where accent I and II were varied systemically in the pragmatic conditions NEWINFO and CONFIRM. One result of the production study (7.2) is that the word accent distinction can be encoded even in connection with an *early fall* in a confirmation, but that optionally, the word accent may be *neutralised* in an *early fall*. It is argued below that this neutralisation provides a further support for the L- hypothesis. The word accent neutralisation is further investigated in the reaction time experiment (7.3).

## 7.2  Production data

The results of the production study presented in this section have also been reported in Ambrazaitis (2007a) and Ambrazaitis (2007b).

## 7.2.1   Method

This study is based on Corpus C, which contains recordings from nine Swedish speakers. As described in more detail in 4.3, the test material of this corpus consists of responses like *Ja, det var med bilen* 'Yes, by car' for accent I and *Ja, det var med bilar* 'Yes, with cars' for accent II. Five different (near minimal) accent I/II pairs were used, and the 10 sentences were elicited in the conditions NEWINFO and CONFIRM. In total, the corpus of this study (Corpus C) contains 180 utterances (2 pragmatic conditions × 2 word accent conditions × 5 pairs of test sentences × 9 speakers). Since one utterance is missing due to a technical problem, the corpus is reduced to 179 utterances.

As in the studies presented in Chapters 5 and 6, a first step in data analysis consisted in a qualitative (auditory and visual) inspection of spectrograms and $F_0$ tracks using Praat (Boersma and Weenink URL). As in Chapter 6, each utterance was categorised according to the $F_0$ pattern associated with the nuclear accented word, as either being realised with a *high+fall* (H-), or with an *early fall* (i.e. the hypothesised L-).

It has already been discussed in connection with the pilot study presented in 6.2, that the pattern labels *high+fall* and *early fall* were coined based on accent I, but are applicable to accent II, as well. Hence, again, a word was classified as being accented with an L- if its stressed syllable is associated with a predominantly falling (accent II) or low (accent I) $F_0$ pattern, and without any further crucial $F_0$ rise that would reflect an H- in the stressed (in the case of accent I) or post-stress (accent II) syllable.

As discussed in more detail in 7.2.2 below, the inspection of the data revealed that the realisation of the word accent distinction in connection with an *early fall* pattern is variable. In order to capture this variation quantitatively, the following $F_0$ measures were taken for the utterances produced with an *early fall* in CONFIRM:

$\overline{F_0pre}$ : average $F_0$ over the last 30 ms of the sonorant portion of the pre-stress syllable

$\overline{F_0V_1}$, $\overline{F_0V_2}$, $\overline{F_0V_3}$ : average $F_0$ over the first, second, and third 33.3% of the stressed vowel

These measures provide an impression of the pitch change from the pre-stress syllable (e.g. [me] in [me ˈbiːlən]) to the stressed vowel (e.g. [iː] in [ˈbiːlən]), as well as during that vowel. All values were expressed in semitones (re 100 Hz). The extraction of the measures was performed semi-automatically. First, some relevant points in time were labelled by hand for each utterance using wave-form diagrams and spectrograms in Praat. These points were the stressed vowel onset

and offset, as well as the offset of the sonorant portion of the pre-stress syllable, as exemplified in [pʰo|st|uː|lən], where the labelled points are indicated by the | sign. In a second step, a Praat script was written in order to extract the measures defined above.

## 7.2.2   Results

In this study, the judgement and classification of the intonation patterns was not always as trivial as in the previous studies. This might be due to the fact that the recordings of Corpus C have a less controlled and artificial character, which is related to the method of elicitation (cf. 4.3). That is, on the one hand, the utterances sound more natural and almost like spontaneous speech. On the other hand, the speakers behaved more variably and for some speakers, the degree of articulatory reduction is rather high. Therefore, the frequencies of occurrence reported for different pattern types or aspects of them in the tables and in the text of this section must be regarded as approximate.

Table 7.1 displays the results of the classification of the recordings as either being accented with an H- (*high+fall*), or with a hypothesised L- accent (*early fall*). While in most cases, it was still easy to decide whether an utterance contained an H- or not, there were some less clear cases, e.g. when a focal accent rise was present but extremely reduced, or when the post-stress syllable in an accent II word was produced with creaky voice. In such cases, where both categories (H- vs. L-) could be perceived, the pattern was counted as H-.

Table 7.1 shows that the NEWINFO utterances were generally produced with a *high+fall*, although four exceptions were found for one speaker. The *early fall* was predominantly used in CONFIRM. Hence, the distribution of pattern types displayed in Table 7.1 differs from the distributions found in the previous chapters. In Chapters 5 and 6, the results showed that an *early fall* was one possible pattern used in CONFIRM, while the *high+fall* was still dominant, in both NEWINFO and CONFIRM.[1] In the present data, however, the *early fall* is in fact the dominant pattern used in CONFIRM by 8 of the speakers, while the *high+fall* is only used exceptionally in CONFIRM. In these cases of H- in a confirmation, however, the H- was produced with a reduced $F_0$ peak, as also observed in Chapters 5 and 6.

Figure 7.1 displays two examples (one for each word accent) of utterances produced with H- on the target word in the NEWINFO condition. The H- is clearly

---

[1]The early fall occurred in 28.9% of the utterances in the CONFIRM condition in Chapter 5; the corresponding figure from Chapter 6 is 24.4% for all three focus locations pooled, or 35.6% if only the final focus is counted, which more closely corresponds to the situations in Chapter 5 and in the present chapter. However, in the pilot study presented in 6.2, the early fall occured in 37.8% (all focus locations), but in 66.7% of the cases for final focus, i.e. in a majority of the cases in a manner similar to the present study.

Table 7.1: Relative frequencies (%) of $F_0$ patterns (Swedish speakers).  Superscript: absolute frequency; number of speakers who produced the pattern in at least one of the five test sentences. Example: $86.7^{39;8}$ = pattern produced at least once by 8 speakers, in 39 of 45 cases (86.7%) in total.  One case is missing for CONFIRM, accent I.

|            |             | $F_0$ pattern type | | |
| Condition  | Word accent | *early fall* | *high+fall* | Total |
|------------|-------------|--------------|-------------|-------|
| NEWINFO    | I           |              | $100^{45;9}$ | 45 |
|            | II          | $8.9^{4;1}$  | $91.1^{41;9}$ | 45 |
| CONFIRM    | I           | $86.4^{38;8}$ | $13.6^{6;2}$ | 44 |
|            | II          | $86.7^{39;8}$ | $13.3^{6;2}$ | 45 |

identified by the $F_0$ peak in the stressed syllable [biː] for accent I, and in the post-stress syllable [laɹ] for accent II. In the examples in 7.1, an additional H- accent was produced on the verb *var* [vɑː] 'was'.

Figure 7.2 displays two typical examples of utterances produced with a hypothesised L- accent (i.e. lacking an H-) in the CONFIRM condition. These confirming responses were generally produced with a hat pattern, with a rise either on the initial *ja* or the word *var* and a fall on the target word. In Chapter 6, a similar pattern was observed for final focus in confirmations, where a rise was produced on the initial word *Wallander* and a fall on final word *November*. This pattern, however, did not result in a hat pattern, which was explained by the fact that the medial word *förlänger* was (non-focally) accented, too.

Several parameters varied in the productions of confirmations, as discussed in what follows and illustrated in Figures 7.2, 7.3, and 7.4. First, a short pause between *ja* and the rest of the utterance was occasionally inserted by 5 speakers, in 19 cases in total (e.g. Figure 7.2a). Second, the concatenation between the initial rise and the final early fall was variable: Often, a high plateau-like contour was found, as in Figure 7.2b. In several cases, however, it is more appropriate to speak of a simple rise (on *ja*) plus fall throughout the utterance, cf. Figure 7.4b. Third, in approximately 20 cases (of 89) in total, or in 10-50% of the cases per speaker (although 0% for one speaker), a word from the intermediate part of the utterance (mostly *var* 'was'), received intonational prominence, as indicated by a separate pitch movement (e.g. Figure 7.2a).

The realisation of the $F_0$ fall in the target word varied also due to the word accent category. For accent I, a low $F_0$ level is usually reached already at the onset of the stressed vowel, rendering a flat low pitch during that vowel, cf. Figures 7.2a
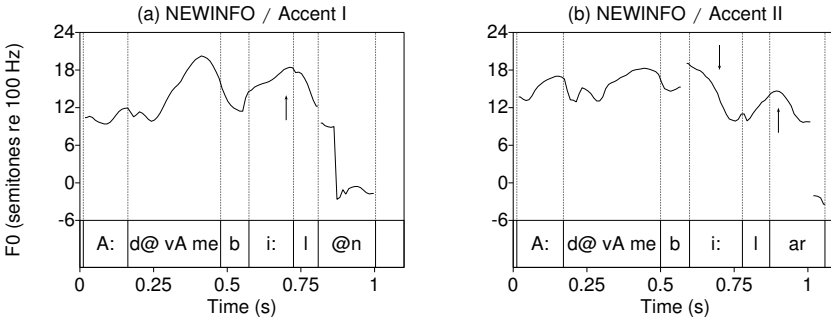
Figure 7.1: $F_0$ courses and SAMPA transcriptions of two examples produced by the female speaker SF6 in NEWINFO. **(a)** *bilen* (accent I) 'the car'; **(b)** *bilar* (accent II) 'cars'. Arrows indicate (a) the focal (H-) rise in the stressed vowel and (b) word accent (H*+L) fall in the stressed vowel and the H- peak, which is located in the post-stress vowel for accent II.
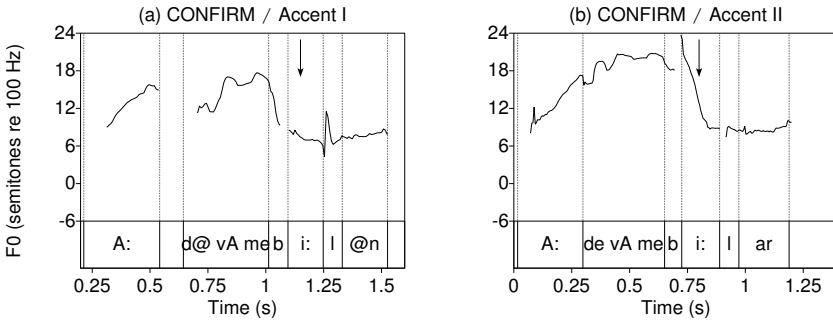


Figure 7.2: $F_0$ courses and SAMPA transcriptions of two examples produced by the female speaker SF7 in CONFIRM. **(a)** *bilen* (accent I) 'the car'; **(b)** *bilar* (accent II) 'cars' ($F_0$ is falling from a high level). Arrows indicate the $F_0$ pattern in the stressed vowel.

and 7.3a. For accent II, it should be expected to find a delay of the fall (H*+L instead of H+L*), where $F_0$ is high at the onset of the stressed vowel and falling during the vowel, as exemplified in Figure 7.2b. A clear accent II realisation of this kind was found in approximately 23 of those 39 cases in which an early fall was produced in an accent II word in a confirmation. However, in approximately 9 cases (of 39), the accent II realisation was found to be very similar to, or even
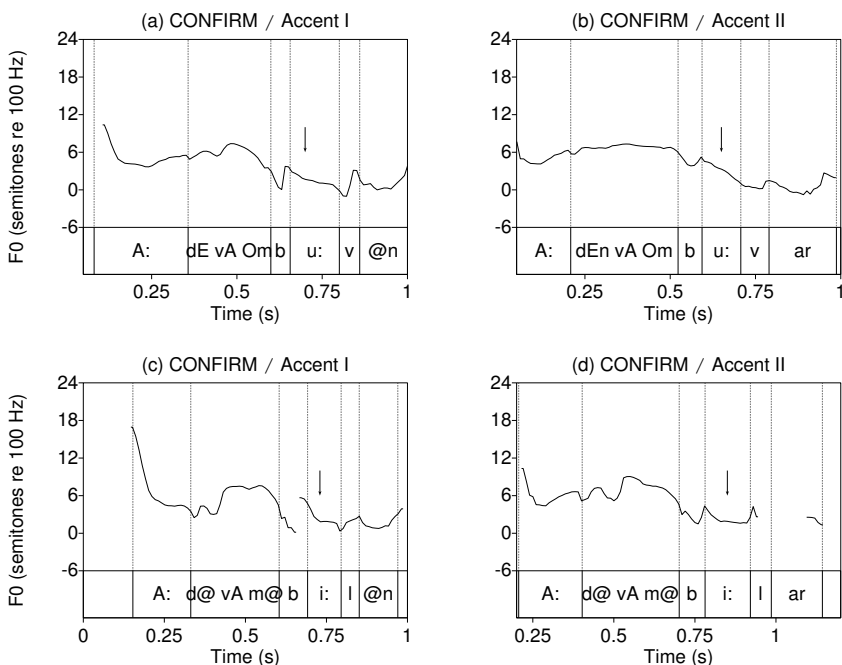
Figure 7.3: $F_0$ courses and SAMPA transcriptions of four examples produced by the male speaker SM2 in CONFIRM. **(a)** *boven* (accent I) 'the villain'; **(b)** *bovar* (accent II) 'villains' ($F_0$ is falling from a medium level); **(c)** *bilen* (accent I) 'the car'; **(d)** *bilar* (accent II) 'cars' ($F_0$ in stressed vowel is comparable to accent I). Arrows indicate the $F_0$ pattern in the stressed vowel.

indistinguishable from a typical accent I pattern, as shown in Figures 7.3d and 7.4. That is, a low pitch is perceived from the vowel onset in these cases. $F_0$ tracks are elusive in some of these 9 cases due to a creaky voice quality (e.g. Figure 7.4a), but perceptually, the vowel is low-pitched. In the remaining 7 cases (of 39), $F_0$ at vowel onset is higher than typically observed for accent I, but lower than the pre-stress $F_0$ level, cf. Figure 7.3b.

That is, some speakers seem to have *neutralised* the word accent distinction in a nuclear *early fall* pattern in a confirmation. In order to express the distinction between accent I and accent II in the *early fall* pattern in quantitative terms, the $F_0$ contours of the relevant 77 cases by 8 speakers (all *early falls* produced in CONFIRM, cf. Table 7.1) were investigated in more detail using the $F_0$ measures defined in 7.2.1.
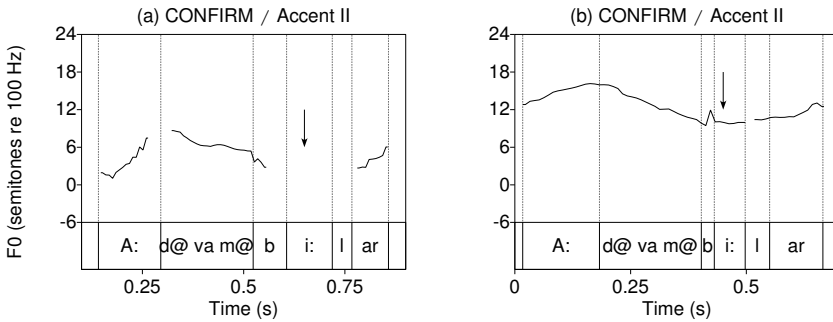
Figure 7.4: $F_0$ courses and SAMPA transcriptions of two further examples of an accent I like realisation of the accent II word *bilar* 'cars', produced in CONFIRM. **(a)** male speaker SM1; **(b)** female speaker SF3. Arrows indicate the $F_0$ pattern in the stressed vowel.

In particular, it was looked at those properties of the $F_0$ contour that, on the basis of the Lund model, should be most relevant for the word accent distinction: In a 'non-focal' condition, we should find a high–low transition from the pre-stress to the stressed syllable for accent I, but not for accent II. Instead, for accent II, the vowel onset should be high, and a fall is expected throughout the vowel.

The difference measure $\overline{F_0 pre} - \overline{F_0 V_1}$ describes the transition from the pre-stress to the onset of the stressed syllable. Table 7.2 displays the mean values for this measure per speaker and word accent condition. Note that the maximum sample size per condition is $n = 5$ (5 target words per condition), and that the actual sample size is often lower, since some values are missing due to the frequent occurrence of creaky voice.

Table 7.2 shows that all mean values for $\overline{F_0 pre} - \overline{F_0 V_1}$ are positive, regardless of the word accent. That is, even for accent II, there is a pitch drop from the pre-stress to the stressed syllable. Nevertheless, for speakers SF6, SF7, SM3, and to a smaller extent for SF5 and SM5, there is a salient difference between the word accents: The step from the pre-stress to the stressed syllable is larger for accent I than for accent II; and for some speakers (SF6, SF7, SM3, SM5), it is in fact below 1 semitone (st) for accent II, indicating that the stressed vowel has a relatively high-pitched onset. However, for two speakers (SM1, SM2) there seems to be no difference in pitch drop between accent I and accent II. For speaker SF3, the data for accent I are insufficient, and those for accent II are very variable (cf. $sd = 5.76$ st). In fact, for SF3 the raw values range from 0.19 st to 13.29 st.

In order to test whether (only) accent II exhibits the predicted fall through the stressed vowel, the fall range within the stressed vowel was calculated by $\overline{F_0 V_1} -$

Table 7.2: Means ($\bar{x}$) and standard deviations ($sd$) for $\overline{F_0pre} - \overline{F_0V_1}$ in semitones, and sample sizes ($n$) for female (f) and male (m) speakers.

| Speaker | Accent I | | | Accent II | | |
|---|---|---|---|---|---|---|
| (sex) | $\bar{x}$ | $sd$ | $n$ | $\bar{x}$ | $sd$ | $n$ |
| SF3 (f) | (6.56) | . | 1 | 5.63 | 5.76 | 4 |
| SF5 (f) | 4.94 | 1.92 | 5 | 2.15 | 1.37 | 5 |
| SF6 (f) | 9.02 | 5.13 | 5 | 0.61 | 2.30 | 5 |
| SF7 (f) | 9.51 | 1.23 | 5 | 0.80 | 1.11 | 5 |
| SM1 (m) | 1.96 | 2.15 | 3 | 2.07 | 0.92 | 3 |
| SM2 (m) | 1.31 | 0.95 | 5 | 2.00 | 1.69 | 5 |
| SM3 (m) | 4.59 | 1.74 | 4 | 0.27 | 1.18 | 5 |
| SM5 (m) | 2.63 | 1.23 | 4 | 0.87 | 1.55 | 4 |

$\overline{F_0V_3}$; the results are displayed in Table 7.3. Generally, $F_0$ is slightly falling during the vowel even for accent I, but (for the speakers with no or few missing values) the mean fall range is larger for accent II than for accent I. However, this mean difference is rather small for some speakers, indicating that there are a number of accent II cases with accent-I-like patterns, supporting the earlier observations.

It should be stressed that the cases that were classified as having an 'accent-I-like pattern' received this judgment due to their perceptually low vowel. As mentioned above, this low pitch was due to, or accompanied by, creaky voice in some of these cases. Since the missing values in the measurements are mainly due to these cases with creaky voice, some of the clearest instances of word accent neutralisation are not represented in the quantitative results. Hence, the measurements do not fully account for the qualitative observation that the word accent contrast was – to a certain degree – neutralised, but as Tables 7.2 and 7.3 show, they nevertheless support the observations.

## 7.2.3   Discussion

As a first general result, the *high+fall* was the preferred pattern in NEWINFO, while the *early fall* was clearly preferred by the speakers in CONFIRM. That is, in this study a more straightforward usage of the early fall in a confirmation was found than in the previous studies. A likely explanation for this difference is the type of the data, related to the method of elicitation (cf. 4.3).

A further result is that the usage of the *early fall* in a confirmation does not

Table 7.3: Means ($\bar{x}$) and standard deviations ($sd$) for $\overline{F_0V_1} - \overline{F_0V_3}$ in semitones, and sample sizes ($n$) for female (f) and male (m) speakers.

| Speaker | Accent I | | | Accent II | | |
|---|---|---|---|---|---|---|
| (sex) | $\bar{x}$ | $sd$ | $n$ | $\bar{x}$ | $sd$ | $n$ |
| SF3 (f) | . | . | 0 | 2.64 | 1.96 | 4 |
| SF5 (f) | 1.30 | 0.72 | 4 | 3.86 | 1.15 | 5 |
| SF6 (f) | 0.94 | 1.54 | 5 | 7.15 | 1.51 | 5 |
| SF7 (f) | 1.49 | 0.71 | 5 | 8.07 | 1.38 | 5 |
| SM1 (m) | (2.30) | . | 1 | (2.30) | . | 1 |
| SM2 (m) | 1.89 | 0.41 | 4 | 2.49 | 0.81 | 5 |
| SM3 (m) | 1.34 | 0.68 | 4 | 5.67 | 1.92 | 5 |
| SM5 (m) | 1.80 | 0.48 | 4 | 2.67 | 0.78 | 4 |

seem to depend on the word accent category of the nuclear word, since the *early fall* occurred equally often in connection with accent I and accent II (Table 7.1). As the examples in Figure 7.2 and the measurements (Tables 7.2 and 7.3) have shown, the word accent distinction can still be maintained in an *early fall*. The difference between accent I and II typically found in the nuclear *early fall* is basically the same as the one found in non-focal (i.e. pre- or post-focal) realisations of the word accents: the HL gesture is timed later in accent II than in accent I, rendering the onset of the stressed vowel high in accent II, but low in accent I.

However, in some cases, an accent II word was produced with an early, accent-I-like timing of the HL in the *early fall* pattern. Thus, the results suggest that the word accent distinction can be *neutralised* in connection with the *early fall*, as produced in a confirmation context, and that the result of the neutralisation is an accent-I-like pattern. It can be stated informally, that these productions indeed sound 'like accent I' on the one hand, but on the other hand, that they do not sound 'wrong'.[2]

Why does this neutralisation occur? This question actually implies two parts: First, why is it acceptable that the neutralisation occurs, and second, what is it that actually triggers this neutralisation? An answer to the first question might seem obvious. As discussed in 3.3.1, the word accent can facilitate the morphological processing, but generally, its functional load is rather low. In a confirmation, where the accented word represents a given discourse referent, it can thus be ar-

---

[2]Some examples of accent II words from the present data, produced as confirmations with an accent-I-like early fall were presented to some native speakers of Swedish.

gued that the information which is contributed by the word accent is less relevant than in a new-information context.The low functional load of the word accent in a given-information context might thus provide a pre-condition for the neutralisation, but it does not fully explain its occurrence. A possible answer to the second partial question is simply that the neutralisation occurs, like any other articulatory reduction, for economic reasons, i.e. as an effect of hypo-speech (Lindblom 1990). This explanation, however, has at least two caveats, which are related to the fact that we are not dealing with a case of *de-accentuation*, but with a (salient) *nuclear* accent gesture.

First, since these cases of *accented* words already imply a certain articulatory cost, the hypo-speech explanation would also need to explain why an accent I-like pattern, i.e. an earlier timing of the fall, implies an articulatory advantage compared to a later timing of the fall. It would seem that the hypo-speech explanation would be more applicable to cases of de-accentuation.[3] Second, the observed neutralisation has occurred in a prominent position in the utterance compared to non-focal (i.e. pre- or post-focal) accents. If the observed word accent neutralisation were, in the first place, a standard hypo-speech effect, then this neutralisation should be expected even more in the least prominent (accented) context, namely in pre- or post-focal accents. But it has typically been reported that the word accent distinction is maintained in such contexts (e.g. Bruce 1977).

An alternative explanation of this neutralisation can be proposed if we assume that the tonal pattern observed in the nuclear *early fall* is not primarily determined by the word accent, but by intonation, more specifically by a *sentence accent*, as suggested in the L- hypothesis. However, for this explanation, to be further developed below, it would seem more adequate to represent this sentence accent as a *falling* pattern rather than a *low* tone (L-) only. Hence, it is suggested here to revise the L- hypothesis and instead to regard the *early fall*, observed as a nuclear accent pattern in the various studies in this thesis, as a *falling sentence accent*. This accent can be represented as H+L- in a Lund model style AM transcription and is thus both functionally and formally comparable to the early peak accent in German (H+L* in GToBI notation).

The (revised) **H+L- hypothesis** thus assumes that in a nuclear *early fall* pattern, the tonal fall represents an intonational gesture. This intonational gesture has basically the same form as a non-focal accent I pattern, and hence, in an accent I word, the two underlying patterns – the word accent H+L* and the sentence ac-

---

[3]Obviously, if a pre- or post-focal word were completely de-accented, this would also imply a neutralisation of the word accent, at least of its tonal correlates. De-accentuation implies that a word is integrated into another prosodic word, something which is, for Swedish, typically only reported in connection with compounds (e.g. Bruce and Hermans 1999), or with function words (cf. the example in Table 2.1). That is, de-accentuation of content words is less usual in Swedish than in languages such as German and English.

cent H+L- – are realised simultaneously, by the same tonal gesture. In an accent II word, then, the two gestures are not entirely congruent, which results in a delay of the tonal fall.

This delay of the tonal fall renders the onset of the stressed syllable perceptually high (H*+L). Thus, an accent II configuration of the H+L- sentence accent might have some perceptual similarity with an accent I configuration of the H- sentence accent, which can also be realised with a high stressed-vowel onset. On the one hand, this similarity can hardly be expected to cause any serious confusions of intonation or word accent patterns in natural communication, since, as discussed in 3.3.1, there are practically no minimal pairs for Swedish word accents. On the other hand, if the word accent, especially accent II, were not encoded in connection with an H+L- intonation, this would imply a low stressed vowel and thus an enhancement of the contrast between an H+L- and an H- pattern. This could imply a support for the perception of the H+L-, and hence a communicative advantage. That is, the H+L- sentence accent might provide a motivation for a word accent reduction. The observed word accent *neutralisation* can thus be explained as a complete *reduction* of the word accent in connection with an H+L- intonation pattern. A word realised with a falling nuclear accent and neutralised word accent is hence *de-accented* in the sense that the *word accent* is elided, but it is still *accented* by means of a *sentence accent*.

To summarise, this study has supported the finding in Chapters 5 and 6 that an *early fall* pattern can be used in Swedish for the expression of confirmation. This falling sentence intonation can be encoded in parallel with the word accent, but the word accent can also be neutralised in this context. A two-step explanation of this word accent neutralisation has been suggested, i.e. an attempt was made to explain both the pre-condition and a motivation for the neutralisation. First, the neutralisation is *acceptable*, because the functional load of the word accents is generally low in Swedish, which is even more the case when the accented discourse referent represents given information, as in a confirmation. Second, the neutralisation is also *motivated*, since it contributes to the enhancement of the intonational contrast between a rising (H-) and a falling (H+L-) sentence accent. The explanation of neutralisation thus also provides a further support for the L-hypothesis, which, however, has been modified into an H+L- hypothesis.

Finally, it must be admitted that the database of this study is rather small. Thus, it might be the case that the observed phenomenon is not representative of Swedish speakers in general, or even that the neutralisation would be perceived as odd by a majority of Swedish speakers. Moreover, this study has only looked at $F_0$. Therefore, we cannot exclude the possibility that the accent-I like accent II realisations still have some accent II quality, related to further parameters beyond $F_0$. The following section presents a perception experiment which attempts to test whether a neutralisation in fact is acceptable in an H+L- intonation, and whether

the proposed explanation of this neutralisation in terms of a contrast enhancement is perceptually plausible.

## 7.3   Perception data

The present section presents a perception experiment, using reaction time measurements, which attempts to test (a) whether a neutralised word accent in an H+L- intonation is perceptually acceptable in a confirmation, and (b) whether the proposed explanation of this neutralisation in terms of a contrast enhancement is perceptually plausible.

For that, the study applies a mismatch paradigm. The idea is thus to create stimuli which contain a canonical mismatch between the word type (accent I vs. accent II word) and the tonal pattern (canonical accent I vs. accent II pattern), both in the context of an H- and an H+L- sentence accent. The hypothesis, as outlined in more detail in 7.3.2 below, is that only some of these canonical mismatches will result in a processing difficulty. In the present study, two circumstances could be expected to cause a processing difficulty. A first possibility is that an unexpected combination of lexical material and the tonal pattern – such as a word with an accent II suffix and a tonal pattern appropriate for accent I – might require an increase in processing effort, as shown by the results of a recent ERP study (Roll 2009, cf. 3.3.1). Second, the processing of a stimulus may be impeded because a certain pitch pattern might be ambiguous in the context of a certain word.

One possibility to measure processing difficulties is by means of reaction times. The underlying assumption is that the perception of a less acceptable, or somehow difficult stimulus will cause a longer reaction time, when the listener is assigned a task which implies that s/he has to react to the stimulus in some way. In this study, the task of the subjects is to identify an utterance as either being produced with an H- or with an H+L- intonation, as explained in more detail in 7.3.1. In general, differences in reaction times for different stimuli are often reflected also in the rate of 'correct' identifications of stimuli (e.g. Harley 2008). Hence, for some of the mismatches created in this study, an increased reaction time, but also a decreased identification rate can be expected. Before discussing the hypotheses of this study in more detail in 7.3.2, the design of the study is outlined in the following Section 7.3.1.

### 7.3.1   Design

Short auditive stimulus utterances are presented to subjects whose task is to judge for each stimulus whether the utterance represents a *bekräftelse* ('confirmation')

or rather a *nyhet* ('new information statement'), thereby indirectly identifying the intonation pattern of the stimulus as either H- (*nyhet*) or H+L- (*bekräftelse*).
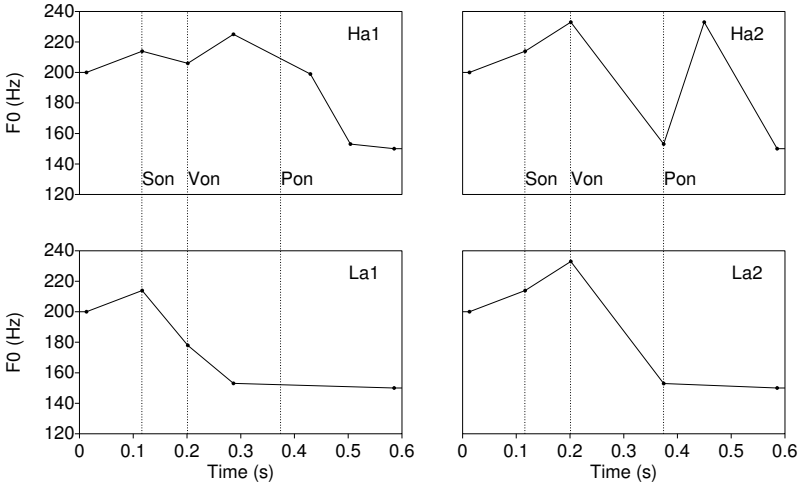


Figure 7.5: The four pitch patterns (Ha1, Ha2, La1, La2) as implemented in the utterance *jag lovar* (cf. Table 7.5) by the female speaker. Three points in time are labelled: stressed-syllable onset ($S_{on}$), stressed-vowel onset ($V_{on}$), and post-stress syllable onset ($P_{on}$). The stimuli are further explained in Section 7.3.3.

The stimuli are short utterances consisting of an unaccented pronoun and an accented verb in present tense, e.g. *Han ljuger* 'He's lying'. The verb form is either an accent I (e.g. *ljuger* 'lie', present tense) or an accent II word (e.g. *frågar* 'ask', present tense). Four different pitch patterns were created and combined with each stimulus phrase, which are illustrated in Figure 7.5: typical *focal accent* (H-) patterns for accent I (Ha1) and accent II words (Ha2), as well as two patterns which were derived from Ha1 and Ha2 by deleting the H- gesture, i.e. by replacing the rising-falling part of the pitch pattern by a fall plus a low (i.e. L) plateau. The resulting patterns are referred to as La1 and La2, and represent the *early fall* pattern (H+L-). That is, the design comprises eight experimental conditions, since the four pitch patterns are combined with accent I words (w1) and accent II words (w2). Labels for the conditions are presented in Table 7.4. Details concerning the stimuli are presented in Section 7.3.3 below.

Both reaction times and the rate of 'correct' identifications of intonation patterns are measured – a rating of a stimulus as *nyhet* or *bekräftelse* is correct, if the stimulus represents an H-, or an H+L- pattern, respectively.

Table 7.4: The eight experimental conditions. Example: **w2Ha1** = test phrase with accent II word, e.g. *jag fattar* (**w2**), combined with a tonal pattern that is canonically appropriate for accent I (H+L*: **a1**), and with an H- sentence accent (**H**).

| | Tonal pattern | | | |
|---|---|---|---|---|
| | H- | | H+L- | |
| Word type | H+L* | H*+L | H+L* | H*+L |
| *-er* (I) | w1Ha1 | w1Ha2 | w1La1 | w1La2 |
| *-ar* (II) | w2Ha1 | w2Ha2 | w2La1 | w2La2 |

## 7.3.2   Hypothesis

The hypotheses for this experiment can be formulated as follows: Four of the eight experimental conditions presented in Table 7.4 comprise a canonically adequate tonal pattern: In w1Ha1 and w1La1, an accent I word is realised with an accent I pattern, and in w2Ha2 and w2La2, an accent II word is realised with an accent II pattern. These four conditions should not provide any processing difficulties and should hence result in relatively short reaction times and high identification rates.

The remaining four conditions represent a canonical mismatch of word type and tonal pattern. However, only two, or possibly three of these are expected to cause a processing difficulty. The hypothesis formulated in 7.2.3 is that an Ha1 and an La2 pattern might have some perceptual similarity. Hence, in a mismatch condition, these pitch patterns might be ambiguous. That is, an Ha1 combined with an accent II word could be mistaken for an H+L- pattern, and an La2 combined with an accent I word as an H- pattern. Increased reaction times and decreased identification rates are thus expected for the conditions w1La2 and w2Ha1.

In contrast, the La1 pattern on the one hand, and the Ha2 on the other hand can hardly be confused with each other or any other pattern, since they exhibit salient tonal characteristics. While the La1 is low-pitched from early in the stressed vowel and through the remainder of the word, the Ha2 comprises a complex falling-rising-falling pitch pattern, with one peak each in the stressed and the post-stress syllable (cf. Figure 7.5). That is, these patterns should result in high identification rates irrespective of a mismatch between word type and tonal pattern. However, it might also be expected that such a mismatch nevertheless causes a processing difficulty, or an increased processing effort. Hence, an increase in reaction time despite a high identification rate could be hypothesised.

However, based on the observation of a word accent neutralisation in con- nection with a nuclear *early fall* in confirmations (7.2.2), it is hypothesised that listeners do not necessarily expect an accent II realisation in an H+L- intonation, and hence an accent I realisation of H+L- in an accent II word should not cause any processing difficulty. An increased reaction time is thus *not* expected for the condition w2La1. A processing difficulty due to the unexpectedness of the tonal pattern is only possibly expected for the condition w1Ha2, i.e. an H- pattern, re- alised as the typical accent two-peaked accent II pattern, in an accent I word.

To summarise, a relative ease in processing, reflected in *short reaction times* and *high identification rates*, is expected for the canonically appropriate condi- tions **w1Ha1**, **w1La1**, **w2Ha2**, and **w2La2**, but also for the canonically inap- propriate condition **w2La1**. A high identification rate should also be obtained for **w1Ha2**, but for this stimulus a processing difficulty, possibly resulting in an increased reaction time, might also be expected. Finally, relatively *long reac- tion times and low identification rates* are expected for the conditions **w2Ha1** and **w1La2**.

### 7.3.3 Method

**Stimuli**

Each experimental condition was represented by 24 stimuli, yielding a total of 8 × 24 = 192 stimuli. For that, 24 stimulus phrases were designed (12 for each word type) and recorded by 2 native speakers of Standard Swedish (1 female, 1 male), yielding 48 base stimulus utterances. As shown in Table 7.5, each stimulus phrase contained a monosyllabic pronoun plus a disyllabic verb. Five different pronouns were used, and each one occurred equally often in an accent I and in an accent II phrase. Furthermore, the set of verbs contained a variety of syllable structures, equally distributed over accent I and accent II words. First, half of the verbs had a long, the other half a short stressed vowel. Second, a subset of the verbs had a voiceless intervocalic consonant (cf. the last two lines for each vowel length category in Table 7.5), while the remaining verbs had a voiced intervocalic consonant or cluster. These variations concerning syllable structure, as well as the variation due to voice (including sex) characteristics of the two stimulus speakers, were not introduced as experimental factors, but rather as a controlled means of raising the external validity of the experiment.

Table 7.5: The 24 stimulus phrases (12 for each word type): Orthographic representations, broad phonetic transcriptions, and approximate idiomatic English translations. All Swedish phrases comprise a monosyllabic pronoun and a disyllabic verb in present tense, with initial stress. Half of the words of each word accent type have a long, the other half a short stressed vowel. Two words per accent type and vowel condition have a voiceless intervocalic consonant. See text for further explanations.

| | Accent I words | | | Accent II words | | |
|---|---|---|---|---|---|---|
| *long vowel* | han lever | [hanˈleːvaɪ] | 'he's alive' | jag lovar | [jaˈloːvaɪ] | 'I promise' |
| | han ljuger | [hanˈjʉːɡaɪ] | 'he's lying' | det haglar | [deˈhɑːɡlaɪ] | 'it's hailing' |
| | hon flyger | [hunˈflyːɡaɪ] | 'she's flying' | jag frågar | [jaˈfɾoːɡaɪ] | 'I'll ask' |
| | jag bjuder | [jaˈbjʉːdaɪ] | 'I'm paying' | vi prövar | [viˈpɾøːvaɪ] | 'we'll try' |
| | det blåser | [deˈbloːsaɪ] | 'it's blowing' | hon snusar | [hunˈsnʉːsaɪ] | 'she takes snuff' |
| | hon röker | [hunˈɾøːkaɪ] | 'she smokes' | han slutar | [hanˈslʉːtaɪ] | 'he's quitting' |
| *short vowel* | jag hinner | [jaˈhinaɪ] | 'I'll manage' | det regnar | [deˈɾɛŋnaɪ] | 'it's raining' |
| | vi säljer | [viˈseljaɪ] | 'we're selling' | vi börjar | [viˈbœɪjaɪ] | 'let's begin' |
| | det dröjer | [deˈdɾœjaɪ] | 'it takes time' | han skojar | [hanˈskojaɪ] | 'he's joking' |
| | det stämmer | [deˈstɛmaɪ] | 'that's right' | hon stannar | [hunˈstanaɪ] | 'she's staying' |
| | jag slipper | [jaˈslipaɪ] | 'I'm excused' | det kvittar | [deˈkvitaɪ] | 'it doesn't matter' |
| | vi sticker | [viˈstikaɪ] | 'we're leaving' | jag fattar | [jaˈfataɪ] | 'I understand' |
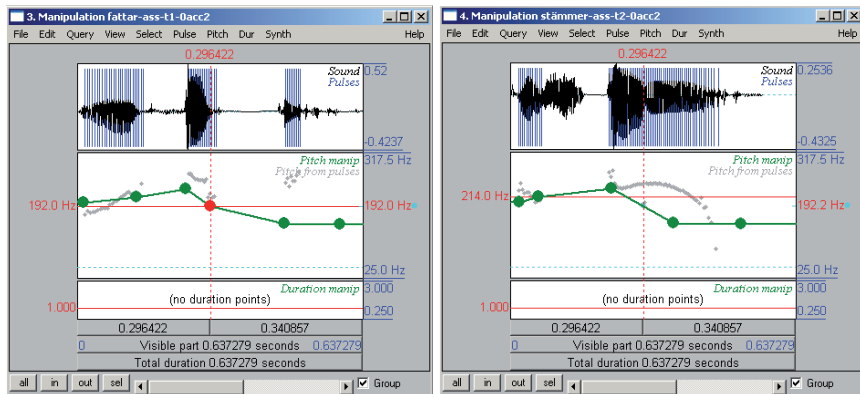
Figure 7.6: Examples of two Praat Windows illustrating implementations of the same pitch pattern (La2) in two different stimulus utterances (by the female speaker): *jag fattar* (left, modelled by 6 $F_0$ points) and *det stämmer* (right, modelled by 5 $F_0$ points). See Table 7.5 for transcriptions and translations.

In each of the 48 recorded utterances the original $F_0$ pattern was replaced by the four pitch patterns introduced in 7.3.1. For that, the recordings were edited and resynthesised using the PSOLA technique (Moulines and Charpentier 1990; Kortekaas and Kohlrausch 1997) in Praat. The pitch patterns were modelled by a small number of 'pitch points' with linear $F_0$ interpolation. The general form including the precise Hz-values of the $F_0$ points of each pitch pattern was kept constant across the 24 phrases per speaker (two different sets of Hz-values were used for the two stimulus speakers, oriented at their original voice). The exact number of points used varied between 5 and 7 and was depending both on the pitch pattern (cf. Figure 7.5) and on the syllable structure of the word (cf. Figure 7.6). No further adjustments were made for words with intervocalic voiceless consonants, i.e. a part of the $F_0$ contour was masked in these words. Furthermore, in all phrases, the $F_0$ points were set at structurally motivated time positions. This implies that the actual distance in ms between two subsequent $F_0$ points for a given pitch pattern varies between the different phrases, which in turn implies that the same pitch pattern is realised with different $F_0$ slopes in the different phrases. However, these variations are counter-balanced between the two word type conditions. Moreover, informal listening confirmed that all 48 versions of each pitch pattern sounded equivalent.

Figure 7.7 displays all four pattern types as implemented in the male-speaker version of *det blåser* (cf. Table 7.5). Due to the very nature of the word accent contrast, being a matter of the *timing* of a pitch fall, the four patterns do not share
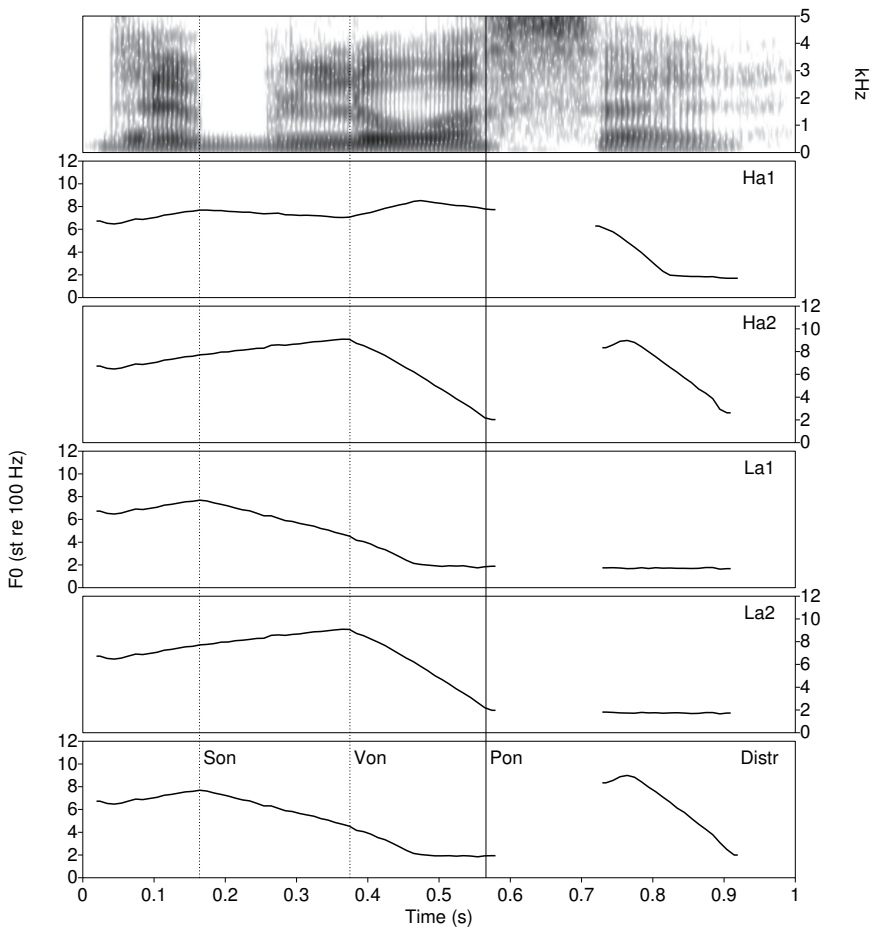
Figure 7.7: Example of all 4 regular pitch patterns (Ha1, Ha2, La1, La2) and the reference pattern used in the distractor stimuli (Dist) as implemented in the utterance *det blåser* (cf. Table 7.5) by the male speaker. The upper panel shows a spectrogram of the utterance as a reference. Three points in time are labelled: stressed-syllable onset ($S_{on}$), stressed-vowel onset ($V_{on}$), and post-stress syllable onset ($P_{on}$). Reaction times are measured from $P_{on}$, cf. text.

a common temporal point of disambiguation. The first disambiguation point is the onset of the stressed syllable ($S_{on}$ in Figure 7.7), where the a1 patterns (La1

and Ha1) diverge from each other as well as from the a2 patterns. The second disambiguation point is the onset of the post-stress vowel ($P_{on}$ in Figure 7.7), until which the a2 patterns (La2 and Ha2) are identical. This could have serious consequences for the interpretation of the results, if, for instance, the a1-patterns generally resulted in shorter reaction times than the a2-patterns, simply due to the accentual timing difference. Since, until slightly after the first disambiguation point, the Ha1 pattern is perceptually more similar to the La2 pattern than to the La1 (both Ha1 and La2 are high-pitched at the onset of the stressed vowel, while La1 is falling), it is assumed that the timing effect is most likely to occur for pattern La1. In order to counter-act this possible effect, a 'reference' pitch pattern was created by adding an accent II-like rising focal gesture to an La1 pattern. This pattern is thus related to both the La1 and the Ha2 pattern as illustrated in Figure 7.7, where the pattern is labelled Distr. It does not reflect any Standard Swedish pitch pattern, but, due to the Ha2-like final pitch movement, it should sound more like an H- than like an L-pattern. This reference pattern was combined with 8 (4 female (2 w1, 2 w2), 4 male (2 w1, 2 w2)) of the 48 recordings, yielding 8 reference stimuli which were used as distractors in the experiment. In order to assure that the subjects would have this pattern in mind while judging a stimulus, the 8 distractor stimuli were distributed over all experimental blocks (cf. below) and, first of all, they were over-represented in the training block (cf. below).

**Subjects and procedure**

The experiment was run on 20 native speakers of Standard Swedish (11 female, aged 19-32, $\bar{x} = 25.6$; 9 male, aged 21-46, $\bar{x} = 31.8$). Two subjects (1 female, 1 male) were left-handed, the remaining right-handed.

The stimuli were presented via headphones, and each stimulus was preceded by a 200 ms noise sound as a start signal; a 500 ms pause was inserted between the noise and the stimulus. The subject's task was to classify the stimulus as either *bekräftelse* 'confirmation' or *nyhet* 'new information' by pressing the *left* or the *right* button of the integrated mouse device on a laptop computer. These two buttons were marked by the letters B and N, written on yellow stickers. The subjects were instructed to use one finger each of both hands for pressing the buttons. For one half of the subjects, the B category was located on the button to the subject's 'strong-hand' side, for the other half on the subject's 'weak-hand' side. The subjects were instructed to respond as quickly as possible. A 1.5 s pause was included after each response before the next trial started again with a noise signal. In case a subject did not respond to a stimulus within a maximum response time set to 5 s, no response was collected for this stimulus and the experiment script jumped to the next trial.

The 200 stimuli (192 regular plus 8 distractors) were divided into 4 experi-

mental lists. Each of the 48 utterances (24 phrases uttered by 2 speakers) occurred once in each list, i.e. each word type occurred 24 times; furthermore, each pitch pattern occurred 12 times. Finally, each list included 2 distractor stimuli. Each list was randomised individually for every subject. Four list orders were prepared, where each list occurred once in each of four positions, in order to compensate for possible effects of list position within the experiment. Two of the four list orders were processed by 6 subjects each, the other two by 4 subjects each.[4] The subjects were given the possibility to take three short breaks, each between two lists, during the experiment.

Prior to the experiment proper, the task – and especially the two response categories *bekräftelse* and *nyhet* – were thoroughly introduced to the subjects by means of examples, including both written examples of short dialogues and sound examples of confirmations and new information statements. The purpose of the sound examples was to establish a clear connection between the two functional categories (*bekräftelse* and *nyhet*) and the two intonation patterns (H+L- and H-). Furthermore, a training block immediately preceded the experiment. The training block consisted of 24 regular plus all 8 distractor stimuli, i.e. 32 stimuli in total. After the training block, as well as on several other occasions during the introduction, the subjects had the possibility to ask questions concerning the procedure.

The software E-Prime 1.2 (Schneider et al. 2002) was used for the experiment. A complete session, including introduction and post-experimental questionnaire, took about 30 minutes per subject.

**Data Analysis**

For each stimulus utterance, the temporal location of the disambiguation point of the pitch patterns (i.e. the post-stress syllable onset $P_{on}$) was identified, and reaction times (RT) were measured from this disambiguation point (cf. $P_{on}$ in Figure 7.7). In order to test for significant differences in mean RT between the conditions, an F1/F2 analysis was performed. That is, two analyses of variance (ANOVA) were calculated, one 'by subjects', and one 'by items'. Hence, for the by-subjects analysis, a mean RT was calculated across all 24 test items for each subject and condition, and the resulting 160 means (20 subjects; 8 conditions) were used as raw data in the ANOVA. For the by-items analysis, a mean RT was calculated across all 20 subjects for each test item and condition, and the resulting 192 means (24 items, 8 conditions) were used as raw data in the ANOVA.

The design contained two factors: WORDTYPE (two levels: w1, w2) and PITCHPATTERN (four levels: Ha1, Ha2, La1, La2). In the by-subjects analysis, both factors are within-subjects factors, while in the by-items analysis, only

---

[4]It was originally intended to have each list order processed by 6 subjects, i.e. to run the experiment with 4 additional subjects.

Table 7.6: Distribution of missing cases over the experimental conditions.

|       | Ha1 | Ha2 | La1 | La2 | Total |
|-------|-----|-----|-----|-----|-------|
| w1    | 4   | 9   | 8   | 15  | 36    |
| w2    | 26  | 8   | 5   | 11  | 50    |
| Total | 30  | 17  | 13  | 26  | 86    |

PITCHPATTERN is a within-items factor, while WORDTYPE is a between-items factor, since half of the 48 test utterances represent accent I (WORDTYPE w1), and the other half accent II (WORDTYPE w2).

The listener judgements of the stimuli as either *bekräftelse* or *nyhet*, were transformed into a measure of identification rate (IR). That is, a stimulus containing an H- or an L- accent was classified as (correctly) identified, if it was judged as a new information statement (H-), or a confirmation (L-), respectively. Also the IR data were analysed by means of an F1/F2 analysis as described above for the RT data. Identification rates in % were used as raw data in the ANOVAs.

### 7.3.4 Results

In 97.76% of the 3840 trials of this experiment (192 stimuli $\times$ 20 subjects), a response was registered successfully, and hence there were 86 missing values (2.24%) in total. Some missing values were expected, since the maximum response time for a trial was set to 5 seconds, as described above. All but five subjects missed a response at least once, and the maximum number of missing values per subject was 14 (7.29% of 192 responses), which applies to two subjects. The missing values are not distributed evenly over the experimental conditions, as shown in Table 7.6. The two conditions which caused most processing difficulties, as discussed below, account for 41 (15+26) of the 86 missing cases in total. The maximum number of missing values per subject and condition was 6 (25% of 24), which occurred in w2Ha1.

Figure 7.8 displays the mean identification rates (IR) and the mean reaction times (RT) for the eight conditions.[5] The results of the by-subjects and the by-items ANOVAs are presented in Table 7.7.

---

[5]In Figure 7.8, only the means of the by-subjects rates (in the case of IR) and the by-subjects means (in the case of RT) are shown. Theoretically, the means of the by-subjects means are identical to the means of the by-items means, if no values are missing. Since some values are missing (cf. Table 7.6), the means in Figure 7.8 would be slightly different if they were based on the by-items means.

Figure 7.8 shows that the identification rates were generally high (between 85% and 90% on average) for six of the eight conditions, but lower for the two conditions w1La2 (61%) and w2Ha1 (70%). That is, an La2 pattern in an accent I word was quite often heard as an H- pattern (i.e. judged as a new-information statement), and an Ha1 pattern in an accent II word was quite often heard as an L-pattern (i.e. judged as a confirmation).
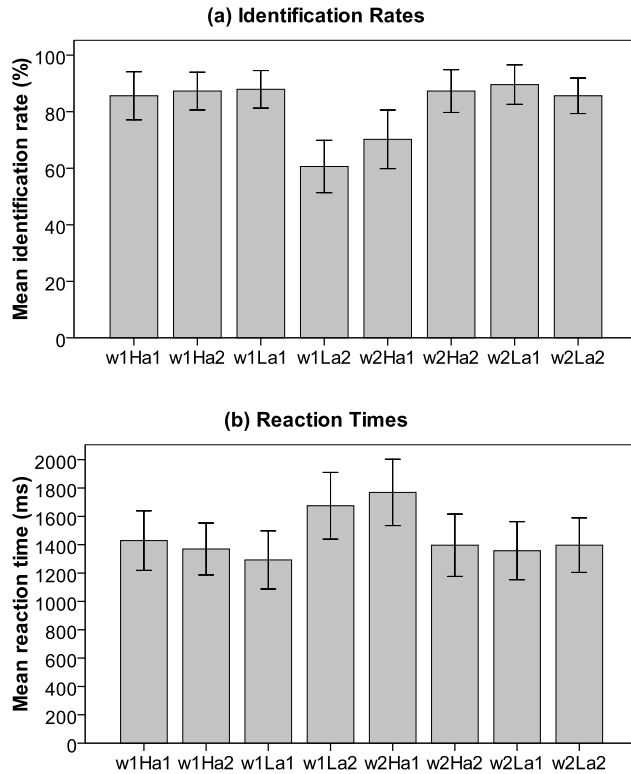


Figure 7.8: (a) Mean identification rates (IR) in % and (b) mean reaction times (RT) in ms for the 8 conditions, based on (n=20) by-subjects rates (IR) and by-subjects means (RT); hence, each bar represents 480 measurements (minus some missing cases), cf. Table 7.6. Error bars represent 95% confidence intervals.

According to the ANOVAs (cf. Table 7.7), both the factor PITCHPATTERN and the interaction of PITCHPATTERN and WORDTYPE have a significant effect on the identification rate. These results confirm the general trends observed in

Table 7.7: Results of the four repeated-measures ANOVAs: degrees of freedom (Greenhouse-Geisser corrected where sphericity cannot be assumed), $F$-values, and $p$-values. Significant results are marked * ($p < .05$) or *** ($p < .001$).

| | | Dependent Variable | |
| Factor | ANOVA | **Identification Rate** | **Reaction Time** |
| --- | --- | --- | --- |
| WORDTYPE | by subjects | $F(1, 19) = 4.74$ $p = .042$ * | $F(1, 19) = 1.79$ $p = .197$ |
| | by items | $F(1, 46) = 2.84$ $p = .099$ | $F(1, 46) = 1.12$ $p = .295$ |
| PITCHPATTERN | by subjects | $F(1.57, 29.90) = 12.38$ $p < .001$ *** | $F(3, 57) = 13.79$ $p < .001$ *** |
| | by items | $F(3, 138) = 33.87$ $p < .001$ *** | $F(2.53, 116.20) = 24.84$ $p < .001$ *** |
| WORDTYPE × PITCHPATTERN | by subjects | $F(2.17, 41.30) = 26.72$ $p < .001$ *** | $F(3, 57) = 16.37$ $p < .001$ *** |
| | by items | $F(3, 138) = 41.82$ $p < .001$ *** | $F(2.53, 116.20) = 24.13$ $p < .001$ *** |

Figure 7.8. That is, different identification rates are obtained for different pitch patterns, but the effect of PITCHPATTERN is different for the two word types. The interaction of PITCHPATTERN and WORDTYPE was resolved by means of an analysis of simple main effects using a Bonferroni adjustment. According to this analysis, for word type I, the mean identification rate is significantly lower for the La2 pattern than for any other pitch pattern[6], while the differences between the rates of La1, Ha1, and Ha2 are not significant. For word type II, only the mean difference between Ha1 and any other pitch pattern is significant.[7]

Table 7.7 also shows that, according to the by-subjects analysis of the identification rates, even the effect of the factor WORDTYPE is significant. According to Figure 7.8, for each word type, the rates for three of four pitch patterns are generally high and similar, and hence, these rates are unlikely to have caused the signif-

---

[6]La2 < Ha1 ($p < .01$), La2 < La1 ($p < .001$), La2 < Ha2 ($p < .001$) for the by-subjects analysis, and La2 < Ha1 ($p < .001$), La2 < La1 ($p < .001$), La2 < Ha2 ($p < .001$) for the by-items analysis

[7]Ha1 < Ha2 ($p < .01$), Ha1 < La1 ($p < .01$), Ha1 < La2 ($p < .05$) for the by-subjects analysis, and Ha1 < Ha2 ($p < .001$), Ha1 < La1 ($p < .001$), Ha1 < La2 ($p < .001$) for the by-items analysis

icant effect of WORDTYPE. However, the rates for the two conditions that were rated relatively low are slightly different: 70% for w2Ha1, and 61% for w1La2. Hence, these figures suggest that the significant effect of WORDTYPE is mainly due to the difference between the rates for w2Ha1 and w1La2. The significant effect of WORDTYPE is, however, not confirmed by the by-items analysis.

Figure 7.8 suggests that the reaction times quite closely correspond to the identification rates. That is, the mean reaction times are relatively short for the conditions with high identification rates, but relatively long for the two conditions with lower identification rate. These results are confirmed by the ANOVAs (cf. Table 7.7). That is, both the effect of PITCHPATTERN and the interaction of PITCHPATTERN and WORDTYPE are significant. However, the effect of WORD-TYPE, which is significant for the identification rate according to the by-subjects analysis (cf. previous paragraph), is not significant for the reaction times. This difference between the RT and the IR data is further discussed in 7.3.5.

Again, the interaction of PITCHPATTERN and WORDTYPE was resolved by means of an analysis of simple main effects using a Bonferroni adjustment. According to this analysis, for word type I, the mean RT is significantly longer for the La2 pattern than for any other pitch pattern[8], while the differences between the mean RTs of La1, Ha1, and Ha2 are not significant. For word type II, only the mean difference between Ha1 and any other pitch pattern is significant.[9]

### 7.3.5   Discussion

The results of this study basically confirm the specific hypotheses concerning the eight experimental conditions (cf. 7.3.2), and hence also the general hypotheses formulated in the introduction to this study, namely (a) that a neutralised word accent in an H+L- intonation indeed is perceptually acceptable, and (b) that the proposed explanation of this neutralisation in terms of a contrast enhancement is perceptually plausible.

Regarding (a), the results have shown that an accent II word realised with an accent I-like *early fall* pattern (La1) did not seem to cause any processing difficulties. It was consistently identified, i.e. classified as a confirmation, and the reaction time was short (cf. the identical results for w2La1 and w1La1). Concerning (b) it has been suggested in 7.2.3 that an accent I configuration of an H-accent (Ha1) and an accent II configuration of an H+L- accent (La2) have some

---

[8]La2 > Ha1 ($p < .05$), La2 > La1 ($p < .001$), La2 > Ha2 ($p < .01$) for the by-subjects analysis, and La2 > Ha1 ($p < .001$), La2 > La1 ($p < .001$), La2 > Ha2 ($p < .001$) for the by-items analysis

[9]Ha1 > Ha2 ($p < .01$), Ha1 > La1 ($p < .01$), Ha1 > La2 ($p < .01$) for the by-subjects analysis, and Ha1 > Ha2 ($p < .001$), Ha1 > La1 ($p < .001$), Ha1 > La2 ($p < .001$) for the by-items analysis

perceptual similarity, related to the high-pitched vowel onset which is expected canonically in accent II. This similarity could possibly motivate a word accent neutralisation in an H+L- intonation, since a neutralised word accent in connection with an H+L- accent would strengthen the low pitch in the stressed syllable and hence enhance the contrast between an H- and an H+L- sentence accent. This hypothesis is supported by the lower identification rates and the longer reaction times found for the mismatch conditions w1La2 and w2Ha1, as compared to the canonically matching conditions w2La2 and w1Ha1.

Finally, the main hypothesis concerning the Ha2 pattern was that it would yield a high identification rate, both in a matching (w2Ha2) and in a mismatch condition (w1Ha2), since it is perceptually salient due to the two-peaked tonal pattern. That is, it is unlikely that this pattern would be confused with an H+L-pattern (i.e. rated as a confirmation), since the H+L- never exhibits a tonal peak on the post-stress syllable. This hypothesis is confirmed, since high identification rates were obtained for the Ha2 pattern.

However, it was argued that a processing difficulty, reflected in longer reaction times, could nevertheless be expected for the mismatch condition w1Ha2, because the pattern could be perceived as unexpected. However, no such processing difficulty is indicated by the results, since the mean reaction time was relatively short. One explanation is simply that the relative ease of classifying the w1Ha2 as a new-information statement could also have resulted in the quick responses. Another possible interpretation of this result is that the Ha2 pattern in connection with an accent I word (w1Ha2) is not perceived as a mismatch, because accent I words are not lexically specified to be associated with a specific tonal pattern. This interpretation would be in line with the traditional view that the word accent distinction in Swedish is privative, or asymmetrical in the sense that only accent II is assumed to be lexically specified (e.g. Riad 1998b, 2006), as also supported by the results of an ERP study presented in Roll (2009), as discussed in 3.3.1.

The results indeed provide some indication that the lexical specification of accent II may have influenced the reaction times in the present study. It has been observed in 7.3.4 that the identification rates for the two ambiguous conditions w1La2 and w2Ha1 are not only lower than for the other six conditions, but that the IR for w1La2 is also lower (61%) than the IR for w2Ha1 (70%). The results of the by-subjects ANOVA showed a significant effect of WORDTYPE, which was argued to be related to the difference between the IRs of w2Ha1 and w1La2. The low IR for w1La2 supports the conclusion drawn above that the processing difficulty associated with w1La2 is primarily due to the fact that the La2 is perceptually similar to the Ha1, and that w1La2 is hence likely to be classified as H-. Such a confusion seems to have occurred less frequently in the w2Ha1 condition, but the reaction times found for w2Ha1 are nevertheless approximately the same (or even slightly longer, cf. Figure 7.8) as for w1La2. This might indicate that

the processing difficulty observed for w2Ha1 is not only explained by the ambiguity of the Ha1 pattern. Hence, the lexical specification of accent II may also have had an influence on the long reaction time found for the mismatch condition w2Ha1. Then, taking into account the short reaction time found for w1Ha2, as discussed above, the present study seems to support the view that accent II is lexically specified, while accent I is not.

However, the asymmetry of the word accent distinction has been treated from two perspectives in the literature (cf. 3.3.1). Besides a difference in the lexical specification of accent I and II, which seems to be supported by the present results, it has also been a widely accepted view that only accent II is associated with a specific (positively specified) tonal pattern, while the realisation of accent I words is entirely determined by utterance prosody. As discussed in 3.3.1, Bruce (1977) and Gussenhoven and Bruce (1999) have assumed another perspective, which also seems to find support from the studies presented in this thesis. We will return to this discussion in the final chapter.

It would seem, however, that a difference in lexical specification between accent I and accent II only had a secondary effect on the results of this study, since the results for the conditions w2La1 and w1La2 should not be expected if lexical specification has been the most influential factor. First, if accent I is not lexically specified, why did the results indicate a difficulty for w1La2? Second, if accent II *is* lexically specified, then why did the results *not* indicate a processing difficulty for w2La1? These two questions have actually already been answered above. To briefly summarise the discussion again, first, w1La2 caused a processing difficulty, because it is ambiguous – it can be perceived as an H+L- (La2) or as an H- (Ha1). Second, w2La1 did not cause any processing difficulty, because the word accent may be neutralised in connection with an H+L- accent. As suggested in 7.2.3 above, the neutralisation implies that the word accent is reduced, and the observed tonal pattern is determined by the H+L- sentence accent. Hence, it would seem that listeners only perceive a mismatch between the expected accent II pattern and a deviating pattern, if this deviating pattern is interpreted as accent I. It could be argued, then, that an La1 pattern is not analysed as an accent I realisation (which would cause a mismatch in a word with an accent II suffix, cf. Roll 2009), but rather as a sentence accent, in line with the H+L- hypothesis.

# Chapter 8

# Summary and conclusions

## 8.1 Background and research questions

The point of departure for the present thesis was the wide-spread assumption that Swedish has a considerably simpler intonation system than German, and the West Germanic languages in general (cf. Gussenhoven's 2004 quotation in Chapter 1). For Swedish, only one type of sentence accent (H-) has been assumed so far, while German has a paradigmatic choice between different sentence accent patterns (e.g. H*, H+L*, L*+H).

On the one hand, languages with tonal contrasts at the word level, like Swedish, can be expected to have a relatively simple system of intonation patterns, at least if the density of lexical tones is high, as in classical tone languages like Mandarin. On the other hand, the lexical usage of tonal contrasts in Swedish is much less intensive and less functionally loaded (cf. 3.3.1). Moreover, as mentioned by Gussenhoven (2004), tone languages can have a more complex intonation system than the one assumed for Swedish, despite their lexical tonal contrasts.

It was hence argued that the difference in intonation patterns represented in contemporary models of German and Swedish intonation does not necessarily reflect a real difference, but is possibly rather a result of different research traditions. In particular, the two basic prosodic functions *prominence* and *grouping* according to Bruce (1998), which interact with the *syntagmatic* dimension of intonation (i.e. the *placement* of sentence accents and boundaries), have traditionally played a primary role in Swedish intonation research. The investigation of *relational* functions of intonation in terms of Barry (1981), which can be related to the *paradigmatic* dimension of intonation (i.e. the *type* of sentence accents and boundaries), is relatively young for the case of Swedish, as compared to the German (or English and Dutch) tradition. However, recent studies, first and foremost the one by House (2005) summarised in 3.3.4, have provided evidence which suggests that

the contemporary Swedish intonation model might need revision.

The overall hypothesis of this thesis is thus that Swedish might have an into-
nation system more similar to German than what has been generally assumed. In
order to test this hypothesis it was decided to compare the intonation patterns, with
a main focus on *nuclear* intonation patterns, produced by German and Swedish
speakers in a variety of pragmatic conditions. These conditions were selected on
the basis of existing hypotheses related to functions of intonation patterns in Ger-
man. The most central pragmatic distinction tested in this thesis is the *focussing*
of *new vs. given* information, where focus is understood as a *semantic* notion as
defined in Krifka (2007) (cf. 2.3.1). A further object of study was the interplay of
the nuclear intonation pattern and word accent realisation in Swedish.

## 8.2   Methods

The research question was approached by means of three production studies and
one perception experiment, using reaction time measurements, as outlined in 4.1.

An important aspect of the research question is how intonation fulfils certain
pragmatic functions.  On the one hand, this calls for the investigation of data
which exhibit a high degree of naturalness, which would best be represented by
spontaneous speech. On the other hand, a high degree of experimental control is
required, because comparable materials need to be investigated for German and
Swedish. In the case of Swedish, word accents must also be controlled for. More-
over, the pragmatic conditions to be investigated are selected in advance. Hence, it
was decided to follow a function-first approach, using constructed materials. The
elicited speech corpora are described in detail in chapter 4.

However, the methods of elicitation were carefully chosen in order to achieve
a certain degree of naturalness and hence ecological validity of the recorded data.
In particular, context descriptions, and in most cases also context questions were
used for the elicitation. Furthermore, a subset of the data (Corpus C) was recorded
by letting subjects act out the prepared test material in pairs.

For the major part of the collected speech data (Corpora A and B), $F_0$ pat-
terns were analysed by means of time and register normalised $F_0$ curves. The
time normalisation method was basically adopted from Yi Xu and colleagues
(e.g. Liu and Xu 2005), while for the $F_0$ domain, the semitone scale was cho-
sen as in, for instance, the FK model (e.g. Fant and Kruckenberg 2006). However,
inspired by Modulation Theory (e.g. Traunmüller 1994), an estimate of the indi-
vidual speaker's $F_0$ baseline was used a reference value.  This method rescales
$F_0$ contours so that a speaker's baseline approximately corresponds to 0 on the
semitone scale. The method thus allows for the calculation of mean contours over
several repetitions and pooled across several speakers, even if different speakers

Table 8.1: A tentative inventory of basic nuclear pattern types of Swedish, compared to German. In Swedish, the tonal pattern comprises an additional word accent gesture (H+L* or H*+L; not shown in the table), preceding or overlapping with the sentence accent. (n) indicates an optional neutralisation of the word accent. See text for further discussion.

|              | German   | Swedish      |
|--------------|----------|--------------|
| *early fall* | H+L* L%  | (n) H+L- L%  |
| *medial fall*| H* L%    | H- L%        |
| *late fall*  | L*+H L%  |              |
| *fall-rise*  | H* LH%   | H- LH%       |
| *simple rise*| L* LH%   | L- LH%       |

are included in different samples (cf. 4.2.3).

## 8.3 Main findings

### 8.3.1 Nuclear pattern types in Swedish

According to the Lund model, Swedish has basically one type of nuclear contour, represented as H- L% (plus a 'continuation rise' (Gussenhoven 2004), i.e. the fall-rise H- LH%), which can be modified phonetically – in terms of parameters like $F_0$ range – in order to fulfil various discourse-related functions (cf. 3.3.3). The results of chapters 5 and 6 have, in part, provided support for this traditional account of Swedish intonation. That is, it was found that the H- L% was indeed used as a kind of default contour in eight of the nine pragmatic conditions of chapter 5, and that the $F_0$ range was varied for different purposes.

Nevertheless, the results of the various studies presented it this thesis support the hypothesis that Swedish has a similar inventory of basic nuclear pattern types as German. In terms of the technical labels defined in chapter 5, this inventory comprises a *simple rise*, a *fall-rise*, and two types of falls: an *early fall*, and a *high+fall*. Table 8.1 displays a tentative inventory of basic nuclear pattern types of Swedish suggested by the results of this thesis, represented in AM-style transcriptions. The table is further discussed in what follows.

The basic pattern types seem to have basically the same functions in the two languages, although their specific usage may vary. For instance, an *early fall* was used in confirmations and in the summarising of an argument, both by German and Swedish speakers. However, Swedish speakers used this pattern even when the

contextually given, nuclear accented word occurred in a question context, where German speakers preferred their *simple rise* pattern. It would thus seem that a basic underlying function of the *early fall* in both Swedish and German is to lend nuclear prominence to a word and at the same time marking it as *given*, but that its pragmatic usage may vary, in line with the proposal by Niebuhr (2007b).

   In order to account for the four nuclear pattern types proposed here for Swedish, it was proposed that a second type of sentence accent – a low L- – should be assumed besides the well established high H-. Furthermore, based on the results of chapter 7, it was suggested for the case of the *early fall* pattern that the proposal of an L- should be revised, substituting the L- by a falling accent H+L-. The thesis did not touch upon the question whether this H+L- is also appropriate for describing the *simple rise*. In analogy with German, it might seem more adequate to assume a three-fold sentence accent paradigm even in Swedish, as indicated by Table 8.1. But since we have not investigated the Swedish *simple rise* in more detail here, and only in connection with accent I, we will leave this issue unresolved. To summarise the essential point of the proposal made on the basis of the present results:

- Swedish has not only one type of sentence accent, but a paradigmatic choice between (at least) two types of sentence accents, like the West Germanic languages.

   A critical point of this proposal is that the assumed falling (or low) *sentence accent*, at a first sight, appears to be difficult to distinguish from a *non-focal word accent*. Hence, one might argue that the Lund model actually can account for the four observed nuclear pattern types, even without assuming an additional sentence accent. However, three main arguments have been discussed in favour of an additional sentence accent, which can be summarised as follows:

1. The observed nuclear patterns of Swedish are formally and *functionally similar* to corresponding patterns in German. Moreover, it would seem that these form–function relations can be explained by means of reference to *biological codes* (Ohala 1983; Ohala 1984; Gussenhoven 2002, cf. 2.4), suggesting that they are of a more universal nature. The assumption of an additional falling (low) sentence accent in Swedish seems to be the most straightforward way of accounting for this functional and formal similarity in Swedish and German.

2. Since the observed nuclear patterns of Swedish have utterance- or discourse related functions, it would seem most appropriate to assume that the accents involved in these nuclear patterns signal *prominence at the utterance level*. Chapter 6 provided some support for this claim since the results suggest that

the assumed falling accent (H+L-) might be used to signal a *narrow focus* on given information. For that, the H+L- has to differ phonetically from adjacent non-focal accentual falls (H+L* for accent I). A (slight) 'boost' of the $F_0$ fall and duration was found for words associated with an assumed H+L- as compared to non-focal (H+L*) words.

3. Finally, the results of chapter 7 suggest that the word accent can be *neutralised* in connection with the proposed H+L- accent. That is, the observed *early falling* pattern in nuclear position is above all determined by intonation, rather than by the word accent.

The results also revealed differences between the German and the Swedish nuclear patterns and their usage. First, some differences in the realisation of the pattern types in the German as compared to the Swedish versions were reported in chapters 5 and 6, which can be explained if we assume that Swedish accent I, in a non-focal context, is realised by means of an $F_0$ fall, as suggested in the Lund model (Bruce 1977). Second, as shown in Table 8.1, the results did not provide any evidence for assuming of a contrast between *medial* and *late falls* in Swedish, a distinction found in German (e.g. Kohler 1991a). Also this difference between the German and the Swedish data might be related to the Swedish word accent contrast.

Moreover, as mentioned above, some Swedish speakers used one contour – the classical H- L% – as a kind of default, and hence made less usage of their various pattern types as compared to German speakers. Furthermore, while there was a general tendency that Swedish and German speakers used the same pattern types in the same contexts, some conditions deviated from this trend (cf. the case of the *early fall* used in a question by Swedish speakers). However, it must also be considered that the test design of the first production study (chapter 5) was based on the available hypotheses on German nuclear pattern types and their prototypical usages. The conclusions drawn here on differences in pattern inventory and usage between German and Swedish must thus be regarded as tentative. Hence, it is possible that further pattern types and also a more intensive usage of different pattern types could be found for Swedish, if a larger and pragmatically more varied material were investigated.

Finally, it should be stressed once more that the present studies have been concerned with the inventory of intonation patterns in Swedish, rather than with their phonological representation. In particular, alternative phonological descriptions of the *early fall* – e.g. in terms of *downstep* – have been discussed for German (e.g. Grice et al. 2009) and might also be considered for Swedish, but this discussion is beyond the scope of this thesis.

### 8.3.2   Pre-nuclear patterns

Although the present thesis focusses on nuclear intonation patterns, the results of Chapter 6 may also have some implications for the modelling of pre-nuclear patterns in Swedish.

A secondary research question was the signalling of a *contrastive topic*. Both German and Swedish speakers have marked a contrastive topic with a pre-nuclear rising accent, and produced an additional nuclear accent on the focus constituent, in line with earlier observations. Both the pre-nuclear and the nuclear accent in the Swedish productions are typically regarded as classical focal accents (H-), the initial one, however, being *timed late*.

It has been proposed previously to classify late-timed initial rises either as variants of the H- accent (Horne et al. 2001; Hansson 2001; Myrberg 2009), or as an initial boundary tone (Roll 2006; Roll et al. 2009). The results of Chapter 6 suggest that both proposals may be adequate, since there seem to be two different types of late-timed initial rises, one of which – the pre-nuclear *accent* – can be used, among other things, for marking a contrastive topic.

Late-timed initial accents are, however, also found in constituents which do not represent a contrastive topic. In the present study, accents on contrastive topics were found to be realised with a *raised* $F_0$ peak as compared to accents on topics for which no contrast was elicited. However, the (lower) H- accents were produced relatively late (i.e. with a temporally broad $F_0$ peak) both in non-contrastive and in contrastive topics. This suggests that the late timing is not a specific feature of a contrastive topic in Swedish, but rather a more general feature of *initial pre-nuclear* H- accents, i.e. such H- accents which are followed by a second, nuclear H- accent later in the phrase. Then, as suggested by Hansson (2001), a general function of this delayed peak in pre-nuclear H- could be the signalling of coherence, in a similar manner as the high plateau in *hat patterns* (cf. Chapter 6).

### 8.3.3   The Swedish word accents

Although this thesis is mainly concerned with intonation, rather than word accents, two different aspects of the results may have implications for the classical debate on the asymmetry of the word accent distinction. As reviewed in 3.3.1, this debate is related to two phenomena (cf. Riad 2009), namely the *phonological representation* on the one hand, and the *lexical specification* (or *markedness*) of the word accents on the other hand. In the traditional view, there is no distinction between these two phenomena: Accent II is regarded as the marked member of the distinction and only accent II is assumed to have a characteristic tonal correlate, while accent I is unmarked and toneless in the sense that its tonal realisation

is exclusively determined by the sentence intonation.

In the alternative account of Bruce (1977), both accent I and II are assumed to contribute a tonal gesture to the surface melodic pattern of an utterance. However, Bruce (as most explicitly expressed in Gussenhoven and Bruce 1999) still admits that accent I and II have an asymmetric relation in the sense that accent I is a default accent. That is, in this account, the lexical specification of the word accents on the one hand, and their phonological representation on the other hand, are considered to be two independent phenomena. The present results seem to be in line with this 'split' account of the word accent distinction, as summarised in what follows.

First, the results of the reaction time experiment presented in Chapter 7 have been argued to be in line with the conception that accent II is lexically specified, while accent I is not.

Second, the studies presented in Chapters 5 and 6 provide some indirect evidence for Bruce's (1977) claim that even accent I has a tonal pattern independent of sentence intonation. In general, the traditional view – in which accent I words are regarded as *unaccented* when not associated with a sentence accent – would be strongly supported if it could be shown that the melodic realisation of assumed 'unaccented' accent I words is entirely determined by the sentence intonation, something which Engstrand (1995) attempted to demonstrate, whose results, however, have been argued to be inconclusive (cf. 3.3.1). One might argue that the support for the traditional view would be persuasive if Swedish non-focal accent I words were realised like non-accented words in related languages like German, which lack lexical pitch accents. However, the data presented in Chapter 6 showed that where German unaccented words were produced with a (high or low) flat $F_0$ pattern, Swedish accent I words were produced with a fall, in line with Bruce (1977).

## 8.4 General conclusions and outlook

The results of this thesis provide a strong support for the hypothesis that Swedish intonation is more complex than what is widely assumed, as e.g. stated by Gussenhoven (2004). In particular, the present studies suggest that the inventory of basic nuclear pattern types in Swedish is quite similar to that of German, comprising at least four, instead of the previously mentioned two or three[1] types of nuclear contours. That is, on the basis of Féry's (2009) typology, which was mentioned in Chapter 1, Swedish and German could both be regarded 'intonation languages'.

---

[1]A three-fold distinction of phrase-final patterns has been suggested by Elert (2000), and also by Hadding-Koch (1961) for South Swedish.

However, the present investigations were based on a rather small selection of pragmatic conditions. That is, the conclusions drawn here should be regarded as preliminary; it might well be the case that further similarities – or differences – between Swedish and German intonation will be discovered in future studies, when further pragmatic conditions are included.

In the elicitation of the data used in Chapters 6 and 7, the idea of *parallel encoding* was applied (e.g. Xu 2009). That is, different types of communicative functions were elicited simultaneously – in Chapter 6: two levels of information structure (narrow focus location and referential status of the focus constituent), and in Chapter 7: the referential status of the nuclear word and its word accent category. Parallel encoding is, however, not only a research paradigm, but rather a research question *per se*, since in order to reach a comprehensive understanding of intonation, one goal must be to learn which types of functions are fulfilled by intonation, which of them can be fulfilled simultaneously, and how they interact in the production of surface $F_0$ contours. The results of this thesis suggest that the parallel encoding approach might be useful also in future research. For instance, in the present thesis, the approach has resulted in the insight that the Swedish word accent distinction can be *neutralised* in connection with a certain sentence accent (H+L-).

An additional task for future studies would be to corroborate the present findings by means of further perception experiments. For instance, the results of Chapter 6 indicate that the assumed H+L- sentence accent can possibly be used to signal a narrow focus in a confirmation, since some acoustic correlates seem to differentiate the H+L- from a word accent H+L*. However, it should, of course, also be tested whether these *acoustic correlates* suffice as *perceptual cues* (cf. Heldner 2001 for a discussion). In this connection it would also seem adequate to investigate further possible acoustic correlates beyond the two taken into account here ($F_0$ and duration), such as spectral emphasis, which might be a further acoustic correlate of – high (H-) as well as falling (H+L-) – focal accents (Heldner 2003). Finally, the additional analysis of spontaneous speech could reveal more insights about the pragmatic usage of the various nuclear intonation patterns assumed for Swedish.

# Bibliography

Adriaens, L. M. H. (1991). *Ein Modell deutscher Intonation: eine experimentell-phonetische Untersuchung nach den perzeptiv relevanten Grundfrequenzänderungen in vorgelesenem Text.* Ph. D. thesis, Technische Universität Eindhoven.

Altmann, H., A. Batliner, and W. Oppenrieder (Eds.) (1989). *Zur Intonation von Modus und Fokus im Deutschen.* Number 234 in Linguistische Arbeiten. Tübingen: Niemeyer.

Ambrazaitis, G. (2005). Between fall and fall-rise: Substance-function relations in German phrase-final intonation contours. *Phonetica 62*(2-4), 196–214.

Ambrazaitis, G. (2007a). Expressing 'confirmation' in Swedish: the interplay of word and utterance prosody. In *Proc. 16th ICPhS*, Saarbrücken, pp. 1093–1096.

Ambrazaitis, G. (2007b). Swedish word accents in a 'confirmation' context. *Proc. Fonetik (Swedish Phonetics Conference), TMH-QPSR 50*(1), 49–52.

Ambrazaitis, G. (2008). On final rises and fall-rises in German and Swedish. In *Proc. Fonetik (Swedish Phonetics Conference)*, Gothenburg, pp. 81–84.

Ambrazaitis, G. (2009a). Focal lengthening in assertions and confirmations. In *Proc. Fonetik (Swedish Phonetics Conference)*, Stockholm, pp. 72–77.

Ambrazaitis, G. (2009b). Swedish and German intonation in confirmations and assertions. In *Nordic Prosody – Proceedings of the Xth Conference*, Helsinki, 2008, pp. 25–36. Peter Lang.

Ambrazaitis, G. and O. Niebuhr (2008). Dip and hat pattern: a phonological contrast of German? In *Proc. Speech Prosody*, Campinas, Brazil, pp. 269–272.

Antoniadis, Z. and H. W. Strube (1981). Untersuchungen zum 'intrinsic pitch' deutscher Vokale. *Phonetica 38*(5-6), 277–290.

Atterer, M. and D. R. Ladd (2004). On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics 32*, 177–197.

Bannert, R. (1984). Towards a model for German prosody. *Working Papers (Dept. of Linguistics and Phonetics, Lund University) 27*, 1–36.

Barry, W. J. (1981). Prosodic functions revisited again! *Phonetica 38*, 120–134.

Baumann, S. (2006). *The Intonation of Givenness – Evidence from German*. Number 508 in Linguistische Arbeiten. Tübingen: Niemeyer.

Baumann, S. and M. Grice (2004). Accenting accessible information. In *Proc. Speech Prosody*, Nara, Japan, pp. 21–24.

Baumann, S. and M. Grice (2006). The intonation of accessibility. *Journal of Pragmatics 38*, 1636–1657.

Baumann, S. and K. Hadelich (2003). Accent type and givenness: An experiment with auditory and visual priming. In *Proc. 15th ICPhS*, Barcelona, pp. 1811–1814.

Beckman, M. E., J. Hirschberg, and S. Shattuck-Hufnagel (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, pp. 9–54. Oxford: Oxford University Press.

Beskow, J., L. Cerrato, B. Granström, D. House, M. Nordstrand, and G. Svanfeldt (2004). The Swedish PF-Star multimodal corpora. In *Proc. LREC Workshop on Models of Human Behaviour for the Specification and Evaluation of Multimodal Input and Output Interfaces*, Lisboa, pp. 34–37.

Beskow, J., B. Granström, and D. House (2006). Visual correlates to prominence in several expressive modes. In *Proc. Interspeech*, Pittsburgh, Pennsylvania, pp. 1272–1275.

Boersma, P. and D. Weenink (URL). *Praat: Doing phonetics by computer*. http://www.fon.hum.uva.nl/praat/.

Bolinger, D. L. (1961). Contrastive accent and contrastive stress. *Language 37*(1), 83–96.

Bolinger, D. L. (1972). Accent is predictable (if you're a mind-reader). *Language 48*, 633–44.

Bolinger, D. L. (1985). Two views of accent. *Journal of Linguistics 21*, 79–123.

Bredvad-Jensen, A.-C. (1981). Tonal interaction between attitude and grammar. In *Nordic Prosody II – Papers from a Symposium*, Trondheim, 1980, pp. 51–62. Tapir.

Bruce, G. (1977). *Swedish Word Accents in Sentence Perspective*. Number 12 in Travaux de l'institut de linguistique de Lund. Gleerup.

Bruce, G. (1981). Tonal and temporal interplay. In *Nordic Prosody II – Papers from a Symposium*, Trondheim, 1980, pp. 63–74. Tapir.

Bruce, G. (1982a). Developing the Swedish intonation model. *Working Papers (Dept. of Linguistics and Phonetics, Lund University) 22*, 51–116.

Bruce, G. (1982b). Textual aspects of prosody in Swedish. *Phonetica 39*, 274–287.

Bruce, G. (1987). How floating is focal accent? In *Nordic Prosody IV – Papers from a Symposium*, Middelfart, 1986, pp. 41–49. Odense University Press.

Bruce, G. (1998). *Allmän och svensk prosodi*. Number 16 in Praktisk Lingvistik. Institutionen för Lingvistik, Lunds Universitet.

Bruce, G. (2005). Intonational prominence in varieties of Swedish revisited. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, pp. 410–429. Oxford: Oxford University Press.

Bruce, G. (2007). Components of a prosodic typology of Swedish intonation. In T. Riad and C. Gussenhoven (Eds.), *Tones and Tunes – Volume 1: Typological Studies in Word and Sentence Prosody*, Number 12-1 in Phonology and Phonetics, pp. 113–146. Berlin; New York: Mouton de Gruyter.

Bruce, G., C.-C. Elert, O. Engstrand, and P. Wretling (1999). Phonetics and phonology of the Swedish dialects – a project presentation and a database demonstrator. In *Proc. 14th ICPhS*, San Francisco, pp. 321–324.

Bruce, G., M. Filipsson, J. Frid, B. Granström, K. Gustafson, M. Horne, and D. House (2000). Modelling of Swedish text and discourse intonation in a speech synthesis framework. In A. Botinis (Ed.), *Intonation. Analysis, Modelling and Technology*, pp. 291–320. Dordrecht: Kluwer.

Bruce, G., J. Frid, B. Granström, K. Gustafson, M. Home, and D. House (1996). The swedish intonation model in interactive perspective. *Proc. FONETIK 96, TMH-QPSR 37*(2), 19–22.

Bruce, G., J. Frid, B. Granström, K. Gustafson, M. Horne, and D. House (1998). Prosodic segmentation and structuring of dialogue. In *Nordic Prosody – Proceedings of the VIIth Conference*, Joensuu, 1996, pp. 63–72. Peter Lang.

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

Bruce, G., B. Granström, K. Gustafson, M. Horne, D. House, and P. Touati (1997). On the analysis of prosody in interaction. In Y. Sagisaka, N. Campbell, and N. Higuchi (Eds.), *Computing Prosody. Computational Models for Processing Spontaneous Speech*, pp. 43–59. Berlin; Heidelberg; New York; Tokyo: Springer.

Bruce, G., B. Granström, K. Gustafson, and D. House (1993a). Interaction of $f_0$ and duration in the perception of prosodic phrasing in Swedish. In *Nordic Prosody VI – Papers from a Symposium*, Stockholm, 1992, pp. 7–22. Almqvist & Wiksell International.

Bruce, G., B. Granström, K. Gustafson, and D. House (1993b). Prosodic modelling of phrasing in Swedish. *Proc. ESCA Workshop on Prosody, Working Papers Dept. of Linguistics and Phonetics, Lund University 41*, 180–183.

Bruce, G. and E. Gårding (1978). A prosodic typology for Swedish dialects. In *Nordic Prosody – Papers from a Symposium*, Lund, 1978, pp. 219–228. Department of Linguistics, Lund University.

Bruce, G. and B. Hermans (1999). Word tone in Germanic languages. In H. van der Hulst (Ed.), *Word Prosodic Systems in the Languages of Europe*, Number 20-4 in Empirical Approaches to Language Typology / EUROTYP, Chapter 9, pp. 605–658. Berlin; New York: Mouton de Gruyter.

Bruce, G., S. Schötz, and B. Granström (2007). Simulekt – modelling swedish regional intonation. *Proc. Fonetik (Swedish Phonetics Conference), TMH-QPSR 50*(1), 121–124.

Bruce, G. and P. Touati (1992). On the analysis of prosody in spontaneous speech with exemplification from Swedish and French. *Speech Communication 11*, 453–458.

Bruce, G., U. Willstedt, and P. Touati (1990). On Swedish interactive prosody: Analysis and synthesis. In *Nordic Prosody V – Papers from a Symposium*, Turku, 1989, pp. 36–48. University of Turku.

Büring, D. (1999). Topic. In P. Bosch and R. van der Sandt (Eds.), *Focus — Linguistic, Cognitive, and Computational Perspectives*, pp. 142–165. Cambridge: Cambridge University Press.

Büring, D. (2007). Intonation, semantics and information structure. In G. Ramchand and C. Reiss (Eds.), *The Oxford Handbook of Linguistic Interfaces*. Oxford: Oxford University Press.

Cambier-Langeveld, T. (2000). *Temporal marking of accents and boundaries*. Ph. D. thesis, Holland Institute of Generative Linguistics.

Cambier-Langeveld, T. and A. E. Turk (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics 27*, 255–280.

Carlson, R. and B. Granström (1973). Word accent, emphatic stress, and syntax in a synthesis rule scheme for Swedish. *STL-QPSR 14*(2-3), 31–36.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and Topic*, pp. 25–55. New York: Academic Press.

Chafe, W. (1994). *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: The University of Chicago Press.

Cohen, A. and J. 't Hart (1967). On the anatomy of intonation. *Lingua 19*, 177–192.

Cooper, W. E., S. J. Eady, and P. R. Mueller (1985). Acoustical aspects of contrastive stress in question-answer contexts. *J. Acoust. Soc. Am. 77*(6), 2142–2156.

Cruttenden, A. (1979). Falls and rises: meanings and universals. In *Proc. 9th ICPhS*, Copenhagen, pp. 371.

Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.

Dombrowski, E. (2003). Semantic features of accent contours: Effects of f0 peak position and f0 time shape. In *Proc. 15th ICPhS*, Barcelona, pp. 1217–1220.

Eefting, W. (1991). The effect of "information value" and "accentuation" on the duration of Dutch words, syllables, and segments. *J. Acoust. Soc. Am. Volume 89*(1), 412–424.

É. Kiss, K. (1998). Identificational focus versus information focus. *Language 74*(2), 245–273.

Elert, C.-C. (1964). *Phonologic Studies of Quantity in Swedish. Based on Material from Stockholm Speakers*. Stockholm: Almqvist & Wiksell.

Elert, C.-C. (1970). *Ljud och ord i svenskan*. Stockholm: Almqvist & Wiksell.

Elert, C.-C. (1972). Tonality in Swedish: Rules and a list of minimal pairs. In E. S. Firchow, K. Grimstad, N. Hasselmo, and W. A. O'Neil (Eds.), *Studies for Einar Haugen*, pp. 151–173. The Hague; Paris: Mouton.

Elert, C.-C. (1984). Phrase-terminal stress in Swedish. In *Nordic Prosody III – Papers from a Symposium*, Umeå, 1983, pp. 43–50. Almqvist & Wiksell International.

Elert, C.-C. (2000). *Allmän och svensk fonetik* (8th ed.). Stockholm: Nordstedts.

Engstrand, O. (1995). Phonetic interpretation of the word accent contrast in Swedish. *Phonetica 52*, 171–179.

Engstrand, O. (1997). Phonetic interpretation of the word accent contrast in Swedish: evidence from spontaneous speech. *Phonetica 54*, 61–75.

von Essen, O. (1964). *Grundzüge der hochdeutschen Satzintonation* (2nd ed.). Ratingen: A. Henn.

Fant, G. and A. Kruckenberg (2004). Prosody by rule in Swedish with language universal implications. In *Proc. Speech Prosody*, Nara, Japan, pp. 405–408.

Fant, G. and A. Kruckenberg (2006). The FK prosody model. Analysis and synthesis. In *Nordic Prosody – Proceedings of the IXth Conference*, Lund, 2004, pp. 77–85. Peter Lang.

Fant, G., A. Kruckenberg, K. Gustafson, and J. Liljencrants (2002). A new approach to intonation analysis and synthesis of Swedish. *Proc. Fonetik (Swedish Phonetics Conference), TMH-QPSR 44*(1), 161–164.

Féry, C. (1993). *German Intonational Patterns*. Number 285 in Linguistische Arbeiten. Tübingen: Niemeyer.

Féry, C. (2009). Indian languages as intonational 'phrase languages'. Manuscript, Universität Potsdam.

Fujisaki, H., M. Ljungqvist, and H. Murata (1993). Analysis and modeling of word accent and sentence intonation in Swedish. In *Proc. Intern. Conf. Acoust. Speech and Signal Processing*, Minneapolis, Minnesota, pp. 211–214.

Gartenberg, R. and C. Panzlaff-Reuter (1991). Production and perception of f0 peak patterns in German. *AIPUK (Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Universität Kiel) 25*, 29–113.

Goldsmith, J. A. (1990). *Autosegmental and Metrical Phonology*. Oxford: Blackwell.

Grabe, E. (1998). *Intonational phonology: English and German*. Ph. D. thesis, Max-Planck-Institut for Psycholinguistics and University of Nijmegen.

Graddol, D. (1986). Discourse specific pitch behaviour. In C. Johns-Lewis (Ed.), *Intonation in Discourse*, pp. 221–237. London; Sidney: Croom Helm.

Grice, M. and S. Baumann (2002). Deutsche Intonation und GToBI. *Linguistische Berichte 191*, 267–298.

Grice, M., S. Baumann, and R. Benzmüller (2005). German intonation in autosegmental-metrical phonology. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, pp. 55–83. Oxford: Oxford University Press.

Grice, M., S. Baumann, and N. Jagdfeld (2009). Tonal association and derived nuclear accents – the case of downstepping contours in German. *Lingua 119*, 881–905.

Gundel, J. K. (1985). "Shared knowledge" and topicality. *Journal of Pragmatics 9*, 83–107.

Gussenhoven, C. (1984). *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.

Gussenhoven, C. (2002). Intonation and interpretation: phonetics and phonology. In *Proc. Speech Prosody*, Aix-en-Provence, pp. 47–57.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Gussenhoven, C. (2007). Notions and subnotions in information structure. In C. Féry, G. Fanselow, and M. Krifka (Eds.), *Basic notions of information structure*, Volume 6 of *Interdisciplinary Studies on Information Structure*, pp. 185–203. Potsdam: Universitätsverlag Potsdam.

Gussenhoven, C. and G. Bruce (1999). Word prosody and intonation. In H. van der Hulst (Ed.), *Word Prosodic Systems in the Languages of Europe*, Number 20-4 in Empirical Approaches to Language Typology / EUROTYP, Chapter 4, pp. 233–271. Berlin; New York: Mouton de Gruyter.

Gårding, E. (1970). Word tones and larynx muscles. *Working Papers (Dept. of Linguistics and Phonetics, Lund University) 3*, 20–46.

Gårding, E. (1977). *The Scandinavian Word Accents*. Number 11 in Travaux de l'institut de linguistique de Lund. Gleerup.

Gårding, E. (1979). Sentence intonation in Swedish. *Phonetica 36*, 207–215.

Gårding, E. (1984a). Chinese and Swedish in a generative model of intonation. In *Nordic Prosody III – Papers from a Symposium*, Umeå, 1983, pp. 79–91. Almqvist & Wiksell International.

Gårding, E. (1984b). Comparing intonation. *Working Papers (Dept. of Linguistics and Phonetics, Lund University) 27*, 75–99.

Gårding, E. (1987). How many intonation models are there in Lund? *Working Papers (Dept. of Linguistics and Phonetics, Lund University) 31*, 1–9.

Gårding, E. (1989). Intonation in Swedish. *Working Papers (Dept. of Linguistics and Phonetics, Lund University) 35*, 63–88.

Gårding, E. and G. Bruce (1981). A presentation of the Lund model for Swedish intonation. In *Nordic Prosody II – Papers from a Symposium*, Trondheim, 1980, pp. 33–39. Tapir.

Gårding, E. and P. Lindblad (1973). Constancy and variation in Swedish word accent patterns. *Working Papers (Dept. of Linguistics and Phonetics, Lund University) 7*, 36–110.

Hadding-Koch, K. (1961). *Acoustico-phonetic studies in the intonation of Southern Swedish*. Number 3 in Travaux de l'institut de phonétique de Lund. Gleerup.

Hadding-Koch, K. and M. Studdert-Kennedy (1964). An experimental study of some intonation contours. *Phonetica 11*, 175–185.

Halliday, M. A. K. (1967a). *Intonation and Grammar in British English*. The Hague; Paris: Mouton.

Halliday, M. A. K. (1967b). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics 3*(2), 199–244.

Hansson, P. (2000). Focally accented topics. In *Proc. Fonetik (Swedish Phonetics Conference)*, Skövde, pp. 65–68.

Hansson, P. (2001). The effect of individual words' information status on accentuation. In *Nordic Prosody – Proceedings of the VIIIth Conference*, Trondheim, 2000, pp. 89–101. Peter Lang.

Hansson, P. (2003). *Prosodic Phrasing in Spontaneous Swedish*. Number 43 in Travaux de l'institut de linguistique de Lund. Department of Linguistics and Phonetics, Lund University.

Harley, T. A. (2008). *The Psychology of Language. From Data to Theory* (3rd ed.). Hove; New York: Psychology Press.

't Hart, J., R. Collier, and A. Cohen (1990). *A Perceptual Study of Intonation. An Experimental-Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.

Haugen, E. (1967). On the rules of Norwegian tonality. *Language 43*(1), 185–202.

Haugen, E. and M. Joos (1972). Tone and intonation in East Norwegian. In D. L. Bolinger (Ed.), *Intonation*, pp. 414–436. Harmondsworth, Middlesex, England: Penguin Education.

Hedberg, N. and J. M. Sosa (2007). The prosody of topic and focus in spontaneous English dialogue. In C. Lee, M. Gordon, and D. Büring (Eds.), *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*, pp. 101–120. Dordrecht: Springer.

Heldner, M. (1997). To what extent is perceived focus determined by f0-cues? In *Proc. Eurospeech*, Rhodes, Greece, pp. 875–877.

Heldner, M. (1998). Is an $f_0$ rise a necessary or a sufficient cue to perceived focus in Swedish? In *Nordic Prosody – Proceedings of the VIIth Conference*, Joensuu, 1996, pp. 109–125. Peter Lang.

Heldner, M. (2001). *Focal accent – F0 movements and beyond*. Ph. D. thesis, Umeå University.

Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics 31*(1), 39–62.

Heldner, M. and E. Strangert (2001). Temporal effects of focus in Swedish. *Journal of Phonetics 29*(3), 329–361.

Hertrich, I. (1991a). The interaction of intonation and accent at the sentence level: Potential perceptual ambiguities. *AIPUK (Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Universität Kiel) 25*, 187–218.

Hertrich, I. (1991b). Perceptual categorisation in terminal intonation patterns of two-accent utterances of German. *AIPUK (Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Universität Kiel) 25*, 295–360.

Hetland, J. and V. Molnár (2001). Informationsstruktur und Reliefgebung. In M. Haspelmath, E. König, W. Oesterreicher, and W. Raible (Eds.), *Language Typology and Language Universals – An International Handbook, Part 1*, Number 20 in Handbooks of Linguistics and Communication Science, pp. 617–633. Berlin; New York: Walter de Gruyter.

Hirst, D. (2005). Form and function in the representation of speech prosody. *Speech Communication 46*, 334–347.

Horne, M. (1991). Why do speakers accent "given" information? In *Proc. Eurospeech*, Genoa, Italy, pp. 1279–1282.

Horne, M., J. Frid, and M. Roll (2006). Timing restrictions on prosodic phrasing. In *Nordic Prosody – Proceedings of the IXth Conference*, Lund, 2004, pp. 117–125. 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*(2), 93–102.

House, D. (1990). *Tonal Perception in Speech*. Number 24 in Travaux de l'institut de linguistique de Lund. Lund University Press.

House, D. (2002). Intonational and visual cues in the perception of interrogative mode in Swedish. In *Proc. ICSLP*, Denver, Colorado, pp. 1957–1960.

House, D. (2003). Perceiving question intonation: the role of pre-focal pause and delayed focal peak. In *Proc. 15th ICPhS*, Barcelona, pp. 755–758.

House, D. (2005). Phrase-final rises as a prosodic feature in wh-questions in Swedish human-machine dialogue. *Speech Communication 46*, 268–283.

House, D. and G. Bruce (1990). Word and focal accents in Swedish from a recognition perspective. In *Nordic Prosody V – Papers from a Symposium*, Turku, 1989, pp. 156–173. University of Turku.

Isačenko, A. V. and H.-J. Schädlich (1970). *A model of standard German intonation*. The Hague; Paris: Mouton.

Jackendoff, R. S. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, Massachusetts: M.I.T. Press.

Jacobs, J. (2001). The dimensions of topic–comment. *Linguistics 39*(4), 641–681.

Jassem, W. (1962). Pitch as a correlate of Swedish word accent. *Dept. for Speech, Music and Hearing Quarterly Progress and Status Report 3*(4), 4–14.

Jassem, W., S. Steffen-Batog, and M. Czajka (1973). Statistical characteristics of short-term average f0 distributions as personal voice features. In W. Jassem (Ed.), *Speech Analysis and Synthesis*, Volume 3, pp. 209–225. Warsaw: Polish Academy of Science.

Johns-Lewis, C. (1986). Prosodic differentiation of discourse modes. In C. Johns-Lewis (Ed.), *Intonation in Discourse*, pp. 199–219. London; Sidney: Croom Helm.

Jones, D. (1969). *An Outline of English Phonetics* (9th ed.). Cambridge: Heffer & Sons.

Kohler, K. J. (1987). Categorical pitch perception. In *Proc. 11th ICPhS*, Tallinn, pp. 331–333.

Kohler, K. J. (1991a). A model of German intonation. *AIPUK (Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Universität Kiel) 25*, 295–360.

Kohler, K. J. (1991b). Terminal intonation patterns in single-accent utterances of German: Phonetics, phonology and semantics. *AIPUK (Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Universität Kiel) 25*, 115–185.

Kohler, K. J. (1995). *Einführung in die Phonetik des Deutschen* (2nd ed.). Berlin: Erich Schmidt.

Kohler, K. J. (1997). Modelling prosody in spontaneous speech. In Y. Sagisaka, N. Campbell, and N. Higuchi (Eds.), *Computing Prosody*, pp. 187–210. Berlin; Heidelberg; New York; Tokyo: Springer.

Kohler, K. J. (2005a). Form and function of non-pitch accents. *AIPUK (Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Universität Kiel) 35a*, 97–123.

Kohler, K. J. (2005b). Pragmatic and attitudinal meanings of pitch patterns in German syntactically marked questions. *AIPUK (Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Universität Kiel) 35a*, 125–142.

Kohler, K. J. (2005c). Timing and communicative functions of pitch contours. *Phonetica 62*, 88–105.

Kohler, K. J. (2006a). Paradigms in experimental prosodic analysis: From measurement to function. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and J. Schließer (Eds.), *Methods in Empirical Prosody Research*, Number 3 in Language, Context, and Cognition, pp. 123–152. Berlin; New York: De Gruyter.

Kohler, K. J. (2006b). What is emphasis and how is it coded? In *Proc. Speech Prosody*, Dresden, pp. 748–751.

Kohler, K. J. and R. Gartenberg (1991). The perception of accents: F0 peak height versus f0 peak position. *AIPUK (Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Universität Kiel) 25*, 219–241.

Kohler, K. J. and O. Niebuhr (2007). The phonetics of emphasis. In *Proc. 16th ICPhS*, Saarbrücken, pp. 2145–2148.

Kortekaas, R. W. L. and A. Kohlrausch (1997). Psychoacoustical evaluation of the pitch-synchronous overlap-and-add speech-waveform manipulation technique using single-formant stimuli. *J. Acoust. Soc. Am. 101*, 2202–2213.

Krifka, M. (2007). Basic notions of information structure. In C. Féry, G. Fanselow, and M. Krifka (Eds.), *The Notions of Information Structure*,

Volume 6 of *Interdisciplinary Studies on Information Structure*, pp. 13–55. Potsdam: Universitätsverlag Potsdam.

Kügler, F. (2007). *The Intonational Phonology of Swabian and Upper Saxon*. Number 515 in Linguistische Arbeiten. Tübingen: Niemeyer.

Kügler, F. (2008). The role of duration as a phonetic correlate of focus. In *Proc. Speech Prosody*, Campinas, Brazil, pp. 591–594.

Ladd, D. R. (1980). *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

Ladd, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge: Cambridge University Press.

Ladd, D. R. and K. E. A. Silverman (1984). Vowel intrinsic pitch in connected speech. *Phonetica 41*, 31–40.

Lahiri, A., A. Wetterlin, and E. Jönsson-Steiner (2005). Lexical specification of tone in North Germanic. *Nordic Journal of Linguistics 28*, 61–96.

Lambrecht, K. (1994). *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*, Volume 71 of *Cambridge studies in linguistics*. Cambridge: Cambridge University Press.

Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Massachusetts: M.I.T. Press.

Lehiste, I. and G. E. Peterson (1961). Some basic considerations in the analysis of intonation. *J. Acoust. Soc. Am. 33*, 419–425.

Liberman, M. and J. Pierrehumbert (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff and R. T. Oehrle (Eds.), *Language Sound Structure*, pp. 157–233. Cambridge, Massachusetts: M.I.T. Press.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H Theory. In W. J. Hardcastle and A. Marchal (Eds.), *Speech Production and Speech Modelling*, pp. 403–439. Dordrecht: Kluwer.

Lindh, J. (2006). Preliminary descriptive f0-statistics for young male speakers. *Proc. Fonetik (Swedish Phonetics Conference), Working Papers Dept. of Linguistics and Phonetics, Lund University 52*, 89–92.

Lindh, J. and A. Eriksson (2007). Robustness of long time measures of fundamental frequency. In *Proc. Interspeech*, Antwerp.

Linell, P. and I. Marková (1993). Acts in discourse: From monological speech acts to dialogical inter-acts. *Journal for the Theory of Social Behaviour 23*(2), 173–195.

Liu, F. and Y. Xu (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica 62*(2-4), 70–87.

Löfqvist, A., T. Baer, N. S. McGarr, and R. Seider Story (1989). The cricothyroid muscle in voicing control. *J. Acoust. Soc. Am. 85*, 1314–1321.

Malmberg, B. (1953). *Sydsvensk ordaccent – en experimentalfonetisk undersökning*, Volume 49 of *Lunds Universitets Årsskrift. N. F. Avd. 1*. Lund: Gleerup.

Malmberg, B. (1959). Bemerkungen zum schwedischen Wortakzent. *Zeitschrift für Phonetik 12*, 193–207.

Malmberg, B. (1966). *Nyare fonetiska rön*. Lund: Gleerups.

Mehlhorn, G. (2002). *Kontrastierte Konstituenten im Russischen. Experimentelle Untersuchungen zur Informationsstruktur*. Europäische Hochschulschriften. Reihe 16: Slavische Sprachen und Literaturen. Frankfurt a.M.: Peter Lang.

Menn, L. and S. Boyce (1982). Fundamental frequency and discourse structure. *Language and Speech 25*, 341–383.

Meyer, E. A. (1937). *Die Intonation im Schwedischen. Erster Teil: Die Sveamundarten*. Stockholm: Fritzes.

Molnár, V. (1991). *Das TOPIK im Deutschen und im Ungarischen*. Number 58 in Lunder Germanistische Forschungen. Almqvist & Wiksell International.

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.

Molnár, V. (2006). On different kinds of contrast. In V. Molnár and S. Winkler (Eds.), *The architecture of focus*, Number 82 in Studies in generative grammar, pp. 197–233. Berlin; New York: Mouton de Gruyter.

Moulines, E. and F. Charpentier (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication 9*, 453–467.

Myrberg, S. (2009). Initiality accents in Central Swedish. In *Nordic Prosody – Proceedings of the Xth Conference*, Helsinki, 2008, pp. 167–178. Peter Lang.

Möbius, B. (1993). *Ein quantitatives Modell der deutschen Intonation. Analyse und Synthese von Grundfrequenzverläufen*. Number 305 in Linguistische Arbeiten. Tübingen: Niemeyer.

Niebuhr, O. (2005). Sequenzen phonologischer Intonationsgipfel. Theoretische Möglichkeiten und empirische Realisierung im segmentellen Kontext. *AIPUK (Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Universität Kiel) 35a*, 55–95.

Niebuhr, O. (2007a). Categorical perception in intonation: a matter of signal dynamics? In *Proc. Interspeech*, Antwerp, pp. 109–112.

Niebuhr, O. (2007b). *Perzeption und kognitive Verarbeitung der Sprechmelodie. Theoretische Grundlagen und empirische Untersuchungen*. Number 7 in Language, Context, and Cognition. Berlin; New York: Walter de Gruyter.

Niebuhr, O. and G. Ambrazaitis (2006). Alignment of medial and late peaks in German spontaneous speech. In *Proc. Speech Prosody*, Dresden, pp. 161–164.

Niebuhr, O. and K. J. Kohler (2004). Perception and cognitive processing of tonal alignment in German. In *Proc. International Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages*, Beijing, pp. 155–158.

Nolan, F. (2003). Intonational equivalence: An experimental evaluation of pitch scales. In *Proc. 15th ICPhS*, Barcelona, pp. 771–774.

O'Connor, J. D. and G. F. Arnold (1973). *Intonation of colloquial English* (2nd ed.). London: Longman.

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica 40*, 1–18.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of $f_0$ of voice. *Phonetica 41*, 1–16.

Peters, B., K. J. Kohler, and T. Wesener (2005). Melodische Satzakzentmuster in prosodischen Phrasen deutscher Spontansprache – Statistische Verteilung und sprachliche Funktion. *AIPUK (Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Universität Kiel) 35a*, 7–54.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. Ph. D. thesis, M.I.T.

Pierrehumbert, J. B. and J. Hirschberg (1990). The meaning of intonation in the interpretation of discourse. In P. Cohen, J. Morgan, and M. Pollack (Eds.),

*Intentions in Communication*, pp. 271–311. Cambridge, Massachusetts: M.I.T. Press.

Pike, K. L. (1945). *The Intonation of American English*. London: Longman.

Prince, E. F. (1981). Toward a taxonomy of given–new information. In P. Cole (Ed.), *Radical Pragmatics*, pp. 223–256. New York: Academic Press.

Prom-on, S., Y. Xu, and B. Thipakorn (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *J. Acoust. Soc. Am. 125*(1), 405–424.

Rathcke, T. and J. Harrington (2006). Is there a distinction between H+!H* and H+L* in standard German? Evidence from an acoustic and auditory analysis. In *Proc. Speech Prosody*, Dresden, pp. 783–786.

Reinholt Petersen, N. (1978). Intrinsic fundamental frequency of Danish vowels. *Journal of Phonetics 6*, 177–189.

Riad, T. (1998a). The origin of Scandinavian tone accents. *Diachronica XV*(1), 63–98.

Riad, T. (1998b). Towards a Scandinavian accent typology. In W. Kehrein and R. Wiese (Eds.), *Phonology and Morphology of the Germanic Languages*, Number 386 in Linguistische Arbeiten, pp. 77–109. Tübingen: Niemeyer.

Riad, T. (2005). Historien om tonaccenten. In C. Falk and L.-O. Delsing (Eds.), *Studier i svensk språkhistoria 8*, Number 63 in Lundastudier i Nordisk Språkvetenskap A, pp. 1–27. Lund: Studentlitteratur.

Riad, T. (2006). Scandinavian accent typology. *Sprachtypol. Univ. Forsch. (STUF) 59*(1), 36–55.

Riad, T. (2009). The morphological status of accent 2 in North Germanic simplex forms. In *Nordic Prosody – Proceedings of the Xth Conference*, Helsinki, 2008, pp. 205–216. Peter Lang.

Rischel, J. (1963). Morphemic tone and word tone in Eastern Norwegian. *Phonetica 10*, 154–164.

Roll, M. (2006). Prosodic cues to the syntactic structure of subordinate clauses in Swedish. In *Nordic Prosody – Proceedings of the IXth Conference*, Lund, 2004, pp. 195–204. Peter Lang.

Roll, M. (2009). *The neurophysiology of grammatical constraints: ERP studies on the influence of prosody and pragmatics on the processing of syntax and morphology in Swedish*. Ph. D. thesis, Lund University.

Roll, M., M. Horne, and M. Lindgren (2009). Left-edge boundary tone and main clause verb effects on syntactic processing in embedded clauses – an ERP study. *Journal of Neurolinguistics 22*, 55–73.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics 1*, 75–116.

Schneider, W., A. Eschmann, and A. Zuccolotto (2002). *E-Prime user's guide*. Pittsburgh, Pennsylvania: Psychology Software Tools.

Selting, M. (1987). Descriptive categories for the auditive analysis of intonation in conversation. *Journal of Pragmatics 11*, 777–791.

Sinclair, J. M. and M. Coulthard (1975). *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. London: Oxford University Press.

Sluijter, A. M. C. and V. J. van Heuven (1996). Spectral balance as an acoustic correlate of linguistic stress. *J. Acoust. Soc. Am. 100*(4), 2471–2485.

Sluijter, A. M. C., V. J. van Heuven, and J. J. A. Pacilly (1997). Spectral balance as a cue in the perception of linguistic stress. *J. Acoust. Soc. Am. 101*(1), 503–513.

Smith, C. S. (2003). *Modes of Discourse – The local structure of texts*. Cambridge: Cambridge University Press.

Strangert, E. and M. Heldner (1994). Prosodic labelling and acoustic data. *Proc. Fonetik (Swedish Phonetics Conference), Working Papers Dept. of Linguistics and Phonetics, Lund University 43*, 120–123.

Terken, J. (1985). *Use and function of accentuation. Some experiments*. Ph. D. thesis, Leijden University.

Terken, J. (1991). Fundamental frequency and perceived prominence. *J. Acoust. Soc. Am. 89*, 1768–1776.

Traunmüller, H. (1994). Conventional, biological, and environmental factors in speech communication: A modulation theory. *Phonetica 51*, 170–183.

Traunmüller, H. (2005). Speech considered as modulated voice. Revised manuscript.

Traunmüller, H. and A. Eriksson (1995a). The frequency range of the voice fundamental in the speech of male and female adults. Manuscript.

Traunmüller, H. and A. Eriksson (1995b). The perceptual evaluation of F0-excursions in speech as evidenced in liveliness estimations. *J. Acoust. Soc. Am. 97*, 1905–1915.

Uhmann, S. (1991). *Fokusphonologie. Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie*. Number 252 in Linguistische Arbeiten. Tübingen: Niemeyer.

Vallduví, E. and E. Engdahl (1996). The linguistic realization of information packaging. *Linguistics 34*, 459–519.

Witting, C. (1977). *Studies in Swedish Generative Phonology*. Uppsala: Almqvist & Wiksell International.

Wunderlich, D. (1988). Der Ton macht die Melodie – zur Phonologie der Intonation des Deutschen. In H. Altmannn (Ed.), *Intonationsforschungen*, Number 200 in Linguistische Arbeiten, pp. 1–40. Tübingen: Niemeyer.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of $f_0$ contours. *Journal of Phonetics 27*, 55–105.

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech communication 46*, 220–251.

Xu, Y. (2009). Timing and coordination in tone and intonation – an articulatory-functional perspective. *Lingua 119*, 906–927.

Xu, Y. (URL). _*TimeNormalizeF0.praat*. http://www.phon.ucl.ac.uk/home/yi/tools.html.

Xu, Y. and X. Sun (2002). Maximum speed of pitch change and how it may relate to speech. *J. Acoust. Soc. Am. 111*, 1399–1413.

Xu, Y. and Q. E. Wang (2001). Pitch targets and their realization: Evidence from Mandarin. *Speech Communication 33*, 319–337.

Xu, Y. and C. X. Xu (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics 33*, 159–197.

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Zonneveld, W., M. Trommelen, M. Jessen, G. Bruce, and K. Árnason (1999). Word-stress in West-Germanic and North-Germanic languages. In H. van der Hulst (Ed.), *Word Prodosic Systems in the Languages of Europe*, Number 20-4 in Empirical Approaches to Language Typology / EUROTYP, Chapter 8, pp. 477–603. Berlin; New York: Mouton de Gruyter.

Öhman, S. (1967). Word and sentence intonation: A quantitative model. *STL-QPSR 8*(2-3), 20–54.

Öhman, S. (1968). A model of word and sentence intonation. *STL-QPSR 9*(2-3), 6–11.

Öhman, S., B. Mårtensson, R. Leanderson, and A. Persson (1967). Cricothyroid and vocalis muscle activity in the production of Swedish tonal accents: A pilot study. *STL-QPSR 8*(2-3), 55–57.

# 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.

44. *Elisabeth Zetterholm*. Voice Imitation. A Phonetic Study of Perceptual Illusions and Acoustic Success. 2003.

45. *Johan Frid*. Lexical and Acoustic Modelling of Swedish Prosody. 2003.

46. *Anastasia Mukhanova Karlsson.* Rhythm and Intonation in Halh Mongolian. 2005.

47. *Susanne Schötz.* Perception, Analysis and Synthesis of Speaker Age. 2006.

48. *Victoria Johansson.* Developmental Aspects of Text Production in Writing and Speech. 2009.

49. *Gilbert Ambrazaitis.* Nuclear Intonation in Swedish. Evidence from Experimental-Phonetic Studies and a Comparison with German. 2009.