Finnish prosody: Studies in intonation and phrasing

Inauguraldissertation
zur Erlangung des Grades einer Doktorin der Philosophie
im Fachbereich Neuere Philologien
der Johann Wolfgang Goethe-Universität
zu Frankfurt am Main

vorgelegt von
Anja Arnhold
aus Berlin

2012
(Einreichungsjahr)

2014
(Erscheinungsjahr)

1. Gutachterin: Prof. Dr. Caroline Féry
(Betreuerin)
2. Gutachter: Prof. Dr. Martti Vainio
Tag der Promotion: 20.06.2013

Acknowledgements

I would like to thank Caroline Féry, who has in all respects been the best superviser I could have imagined, and Martti Vainio, who very kindly welcomed me to Helsinki and supported me in many aspects of my research.

I have been fortunate to conduct some of the research presented here in the inspiring and supportive atmosphere at the University of Potsdam and later at the University of Helsinki. Susanne Genzel, Anja Gollrad, Markus Greif, Shin Ishihara, Gerrit Kentner, Frank Kügler, Anja Kleemann-Krämer Ruben van de Vijver and Thomas Wiederhold have been a big part in making my time in Potsdam pleasant and productive. My colleagues in Vironkatu, Juhani Järvikivi, Heini Kallio, Satu Saalasti, Inkeri Salmenlinna, Antti Suni and Martti Vainio helped me to settle in Helsinki and understand Finnish and Finland a bit better.

In particular, I am grateful to Anna-Kaisa Hiltunen, Juhani Järvikivi, Heini Kallio, Outi Kautonen, Satu Saalasti, Inkeri Salmenlinna and Martti Vainio for informal recordings, discussions sharing native speaker intuitions and checking experimental items. Heini Kallio addionally spoke the prompts for the experiments discussed in Chapters 2 and 5 and helped me edit the textgrids for Chapter 2. Thanks are also due to Mikko Kurimo and Janne Pylkkönen for the original automatic segmentations of these recordings. Antti Suni wrote a script automatically adjusting the fit of these segmentations before the final manual corrections. I am also thankful to him for for synthesising the prompt questions for the experiment in Chapter 4 in a very short time. Juhani Järvikivi, Heini Kallio, Kati Palo and Martti Vainio kindly participated in the pilot perception study discussed in section 3.3. I also thank Werner Arnhold for writing a python script creating LaTeX files that I used to obtain the example pitch contour pictures.

As a scientific contribution and as a readable text, this thesis has profited immensely from comments on earlier versions of different chapters, for which I am very grateful to Caroline Féry, Juhani Järvikivi, Martti Vainio and Kari Suomi; I would especially like to thank Caroline for her thorough and constructive review.

While working on this project, received financial support through grants from the German National Academic Foundation (Studienstiftung des deutschen Volkes, April 2009 to July 2011) and the Anna Ruths-Stiftung (October 2011 to March 2012). From October 2007 to March 2009, I worked as researcher in project FA 255/5 'Morphosyntax und Phonologie von diskontinuierlichen Nominal- und Präpositionalphrasen', financed by the German Research Foundation (DFG). I am thankful to the principal investigators Gisbert Fanselow and Caroline Féry for giving me this opportunity. I am also grateful to my family for financial and emotional support, and to all my friends who have expressed interest and encouragement.

And then, I want to thank Juhani, who has helped me in finishing this dissertation and being happy.

Contents

1 Introduction and Overview **1**

1.1 Introduction . . . 1
1.2 Theoretical framework . . . 2
1.2.1 The autosegmental-metrical framework . . . 3
1.2.2 Information structure . . . 5
1.3 Finnish segmental inventory . . . 7
1.4 Finnish prosodic hierarchy . . . 10
1.4.1 Mora and syllable . . . 10
1.4.2 Foot and prosodic word . . . 12
1.4.3 Phrase-level prosody . . . 15
1.5 Prosodic marking of lexical quantity in Finnish . . . 17
1.6 Prosodic marking of information structure in Finnish . . . 18
1.7 Thesis overview . . . 19

2 Production Experiment I: The influence of information structure, quantity and position **23**

2.1 Introduction . . . 23
2.2 Methods . . . 25
2.2.1 Materials . . . 25
2.2.2 Participants . . . 27
2.2.3 Recording procedure . . . 28
2.2.4 Data editing, measurements and analysis . . . 28
2.3 Hypotheses . . . 29
2.3.1 Tonal contour . . . 31
2.3.2 Pitch scaling . . . 34
2.3.3 Duration . . . 35
2.3.4 Alignment . . . 35
2.3.5 Voice quality . . . 37
2.3.6 Occurrence of pauses . . . 37
2.4 Results . . . 37
2.4.1 Tonal contour . . . 37
2.4.2 Pitch scaling . . . 45

2.4.3 Duration . . . 54
2.4.4 Alignment . . . 58
2.4.5 Voice quality . . . 65
2.4.6 Occurrence of pauses . . . 70
2.5 Discussion and summary . . . 73
2.5.1 Effects of position . . . 74
2.5.2 Effects of focus condition . . . 79
2.5.3 Effects of quantity . . . 86
2.5.4 An alternative account: the nuclear accent hypothesis 90
2.6 Summary . . . 94

3 Refuting the nuclear accent hypothesis 95
3.1 Introduction . . . 95
3.2 Peak alignment in the data from Production Experiment I . . 96
3.2.1 Data and data editing . . . 96
3.2.2 Hypotheses . . . 102
3.2.3 Results . . . 107
3.2.4 Discussion . . . 119
3.3 A pilot study on the perception of peak height and peak timing 125
3.3.1 Data and analysis . . . 125
3.3.2 Results . . . 127
3.3.3 Discussion . . . 129
3.4 Stress and intensity . . . 130
3.4.1 Data and analysis . . . 130
3.4.2 Results . . . 132
3.4.3 Discussion . . . 135
3.5 Conclusion . . . 139

4 Production Experiment II: Syllable structure and information structure 141
4.1 Introduction . . . 141
4.2 Methods . . . 142
4.2.1 Materials . . . 142
4.2.2 Recording procedure . . . 143
4.2.3 Participants . . . 145
4.2.4 Editing and analysis . . . 145
4.3 Hypotheses . . . 146
4.3.1 Alignment . . . 146
4.3.2 Duration, pauses, voice quality, pitch movements exceeding the target words and pitch scaling . . . 148
4.4 Results . . . 149
4.4.1 Duration . . . 150
4.4.2 Pauses and voice quality . . . 152
4.4.3 Pitch movements exceeding the target words . . . 157

4.4.4 Alignment . . . 159
4.4.4.1 Peak alignment . . . 163
4.4.4.2 Alignment of the first f_0 minimum . . . 170
4.4.4.3 Alignment of the second f_0 minimum . . . 170
4.4.5 Pitch scaling . . . 178
4.5 Discussion . . . 182
4.5.1 Tonal analysis and phrasing . . . 182
4.5.2 Focus marking . . . 185
4.5.3 Quantity and syllable structure . . . 186
4.6 Summary . . . 188

5 Perception Experiment: The role of the pitch contour in the perception of compounds and noun phrases 189
5.1 Introduction . . . 189
5.2 Methods . . . 193
5.2.1 Target and filler items . . . 193
5.2.2 Auditory stimuli . . . 194
5.2.3 Procedure . . . 198
5.2.4 Participants . . . 199
5.2.5 Data editing and analysis . . . 199
5.3 Hypotheses . . . 201
5.4 Results . . . 205
5.4.1 Choice . . . 205
5.4.2 Reaction times . . . 218
5.5 Summary and discussion . . . 227

6 Conclusion 232
6.1 Summary . . . 232
6.2 Peak timing . . . 233
6.3 Finnish as a phrase language . . . 235

7 References 238

A Background information on participants of Production Experiment I 248

B Vowel sequences and diphthongs in the materials of Production Experiment I 252

C Treatment of very small pitch movements in the materials of Production Experiment I 254

D Segmental alignment in the materials of Production Experiment I 256

E R code used for editing the data set from production experiment I before the statistical analysis of peak alignment 259
E.1 Determining the relevant L values 260
E.1.1 Minima preceding the peaks 260
E.1.2 Minima following the peaks 262
E.2 Evaluating the measured f_0 maxima 264

F Peak alignment for data sets edited with alternative criteria for perceptually relevant pitch movements 269

G Random effects for models of peak alignment 271

H List of experimental sentences used in Production Experiment II 276

I Background information on participants of Production Experiment II 347

J Models of peak alignment for subsets of data from Production Experiment II 352
J.1 Peak alignment in disyllabic and trisyllabic words 352
J.2 Peak alignment in first syllable contrast subsets 353

K Background information on participants of the perception experiment 360

L List of target and filler items of the perception experiment 364

M Abstract 369

N Deutsche Zusammenfassung 371

O Lebenslauf 373

Chapter 1

Introduction and Overview

1.1 Introduction

Finnish is a Uralic language and differs in several respects from the majority of European languages, which have constituted the point of departure for the development of most linguistic theories. Its best known characteristics probably are its lexical quantity distinction, vowel harmony and an inflectional system with more than ten regularly used nominal cases (for a general grammatical overview see Sulkala and Karjalainen, 1992; Karlsson, 1999). In spite of these differences, descriptions of Finnish intonation are strongly reminiscent of accounts of Germanic languages: Finnish phonologists and phoneticians agree that Finnish intonation is shaped by accents associated with stressed syllables and used to express pragmatic as well as affective meanings (e.g. Iivonen, 1987).

Based on this very similar overall account, one could conclude that within the traditional typological divide into intonation languages, tone languages and lexical pitch accent languages (or stress accent, tone and non-stress accent languages), Finnish should be grouped with English, German and Dutch and is thus an intonation language (for a summary of the traditional typological categories, see Hyman, 2006, where Finnish is also mentioned as belonging to one class with English and Russian, +stress accent and −tone languages). Indeed, Ylitalo (2009) states that in Finnish, "accent is realised in the same way as in, according to Beckman's (1986) classification, stress accent languages" (2009, p. 17).

However, a closer look at the literature provides reasons to doubt that grouping it together with the Germanic languages would appropriately capture Finnish intonation. The pitch contours of intonation languages are commonly characterised by an inventory of accents specifying their differences in shape and alignment as well as their functions and the contexts in which they occur (e.g. Pierrehumbert 1980 for English, Féry 1993 and Grice et al. 2005 for German and Gussenhoven 2005 for Dutch). Such an inventory has not been proposed for Finnish. What is more, Välimaa-Blum (1988, p. 88) has suggested that an inventory containing only one accent "is sufficient for the expression of the basic intonational structure" (p. 114) of Finnish. This statement by and large reflects a prevalent assumption in accounts of Finnish intonation.

The aim of this thesis is therefore to explore Finnish prosody in light of this obvious contrast to better-studied languages. It includes several studies designed to investigate whether an accent inventory can be established for Finnish or whether basic intonational patterns can indeed be accounted for by a single tonal specification. Based on these studies, the thesis also addresses the question how Finnish prosody should be described in a typological perspective. In a nutshell, it finds no basis for establishing an accent inventory and concludes that Finnish is indeed substantially different from languages like English, German and Dutch. The central suggestion of this thesis is therefore that Finnish belongs to a different language type prosodically—a phrase language in the sense of Féry (2010)—and that Finnish prosody is best accounted for not by a pitch accent inventory containing only one accent, but in terms of phrasing. Thus, pitch movements that have traditionally been described as accents are analysed as realisations of phrase-level tones.

Before providing a short overview of this hypothesis and the data it is based on in section 1.7, the remainder of this chapter will establish some background. The following section briefly makes the main theoretical assumptions of this thesis explicit, introducing the autosegmental-metrical model of intonation as the framework of analysis and stating a few basic assumptions regarding information structure. Section 1.3 introduces the segmental inventory of Finnish, before section 1.4 sketches the prosodic hierarchy assumed for Finnish in this thesis and compares it with the existing literature. Finally, sections 1.5 and 1.6 summarise the previous literature on the prosodic marking of lexical quantity and information structure in Finnish to provide a backdrop for the findings presented in subsequent chapters.

1.2 Theoretical framework

This section states the theoretical context of the current description of Finnish prosody. Its aim is to clarify the background and the terminology and assumptions that follow from it, and thus enable the reader's interpretation of the results. Therefore, section 1.2.1 shortly introduces the model employed in the current description of Finnish prosody, while section 1.2.2 sketches the assumptions with respect to information structure.

1.2.1 The autosegmental-metrical framework

The this account of Finnish prosody will be formulated within the autosegmental-metrical (AM) framework (Bruce, 1977; Pierrehumbert, 1980; Ladd, 1996), which has been widely adopted and applied to describe the prosody of typologically very different languages (Gussenhoven, 2004; Jun, 2005b). It models f_0 contours as strings of underlying tonal targets, which are either high (H) or low (L), while a mid level (M) is occasionally also assumed. The framework is autosegmental in that it represents tones and segmental text on separate levels (tiers). The two tiers are then matched by rules associating tones with text, a concept originally developed to account for African tone languages and later applied to English and other European and non-European languages (Leben, 1971; Williams, 1976; Goldsmith, 1976). Thereby, tones can be associated with text one-to-one, one tone can be associated with more than one tone-bearing unit, one unit can be associated with more than one tone or underlyingly present tones can be left unassociated (see the overview in Gussenhoven, 2004, pp. 28–32). The phonological tones are then phonetically implemented according to certain principles. Pierrehumbert (1980) gives a very concrete quantified account arguing that tonal targets are scaled relative to the speaker's baseline and, crucially, to the locally preceding tones. Thus, an H tone late in the sentence is often realised with lower absolute f_0 than an earlier L target. Equally importantly for the reduced theoretical modelling of continuous f_0 contours, the stretches between tonal targets are assumed to be determined by interpolation.

The second crucial aspect of the AM-framework is the representation of prominence relations, which have been modelled in terms of a metrical grid or tree structure (Liberman, 1975; Liberman and Prince, 1977; Hayes, 1980; Selkirk, 1980), and metrical structure as hierarchically ordered domains of phonological processes (Selkirk, 1981; Nespor and Vogel, 1986). Selkirk (1981) refers to the units syllable, foot, prosodic word, phonological phrase, intonational phrase and utterance. Nespor and Vogel (1986) assume the same hierarchy, with the addition of the clitic group ranking between prosodic word and phonological phrase. Subsequent literature has seen a debate about the number of higher prosodic domains needed, especially those intervening between the word and intonation(al) phrase level, which has additionally been complicated by the suggestion of several different labels such as 'intermediate phrase', 'accentual phrase', 'minor phrase' and 'major phrase'. Here, the phonological phrase is called prosodic phrase (p-phrase) and no intermediate levels between it and the intonation phrase (i-phrase) are assumed (see 1). An uncontroversial extension of the original proposal is that tones are associated with moras in several languages, so that this level needs to be included at the bottom of the hierarchy (e.g. Pierrehumbert and Beckman, 1988, for Japanese). Prosodic phrasing is influenced by syntactic phrasing, among other factors, but prosodic and syntactic constituents are frequently not co-extensive (see e.g. Nespor and Vogel, 1986; Selkirk, 2000; Féry and Ishihara, 2009; Féry, 2011). A further factor impacting prosodic phrasing that is relevant in this thesis is information structure (see e.g. Selkirk, 2000; Gussenhoven, 2004; Féry and Ishihara, 2009; Féry, 2011).

(1) The prosodic hierarchy

Utterance (U)
|
Intonation phrase (i-phrase)
|
Prosodic phrase (p-phrase)
|
Prosodic word (PW)
|
Foot (Ft)
|
Syllable (σ)
|
Mora (μ)

With reference to the metrical aspect, Pierrehumbert (1980) proposed that intonational tones in English can be classified as either pitch accents, boundary tones or phrase accents. Pitch accents are tones that are located at metrically strong, accented syllables. They comprise at least one tone, marked with an asterisk, that is associated with the accented syllable. Optionally, they can consist of additional leading or trailing tones, which are not associated themselves, but are realised at a close distance to the starred tone. While pitch accent tones are thus associated with metrically prominent parts of utterances, boundary tones are associated with the beginning or end of a prosodic phrase. Pierrehumbert (1980) originally only conceived of boundary tones associated with the edges of intonation phrases, which she marked by a % sign. To explain systematic pitch movements between the last pitch accent and the boundary tone of an intonation phrase, she introduced phrase accents marked by a minus sign, e.g. L-.[1] Whether the phrase accent should be granted a special status distinguishing it from pitch accents and boundary tones has been the topic of much debate (for a short overview, see Gussenhoven, 2004, pp. 139–141). Crucially, Beckman and Pierrehumbert (1986) reanalysed it as a boundary tone of the intermediate phrase, and the notion that other prosodic domains than the intonation phrase can be marked with boundary tones is widely adopted nowadays. Apart from

[1] In fact, Pierrehumbert's (1980) original notation also attached the - sign to leading and trailing tones of pitch accents. This has however become uncommon even in the ToBI tradition, which continues to use the minus sign for marking phrase accents.

that, Pierrehumbert's original classification is usually considered sufficient to account for the intonation of very different languages (Jun, 2005b).[2]

A terminological decision consistent with adopting the AM-framework is that this dissertation will strictly distinguish between stress and accent. It thereby views stress as a lexically specified property of a syllable which constitutes a potential for a phonetic realisation as prominent (see Lehiste, 1970, p. 150). The term (pitch) accent will be used in the sense of Pierrehumbert (1980) here. That is, it is understood to signify a unit of one or more tones associated with a stressed or otherwise lexically determined prominent tone-bearing unit. In line with this understanding, the dissertation will not—unless when citing previous literature—employ the terms 'word stress' and 'sentence stress', which have frequently been used to designate what will be referred to as 'stress' and (post-lexical) 'accent' here.

1.2.2 Information structure

Another area that will figure prominently in the present work is information structure (for a concise discussion and definition of the core notions of information structure see Krifka, 2007, and the other articles in the same volume). The term describes what Chafe (1976) has called 'information packaging', i.e. the way in which the propositional content of an utterance is organised with respect to the context and the common ground established between the speaker and the listener (or the writer and the (assumed) reader). In this way, utterances are structured and divided into parts according to their information status. Three major divisions have been advocated to be central: topic vs. comment, given vs. new and focus vs. background. The first characterises the fact that some part of the utterance can be described as referring to 'what the utterance is about' (topic), while the other part is the information that is given about it (comment) (see Reinhart, 1981). This aspect will not figure prominently in the thesis, which made no special attempt to elicit topics. The distinction of given vs. new refers to whether the referent of a linguistic expression is treated as something that can be taken for granted or whether the reference has to be newly established in the discourse context at hand—it has in fact been argued to constitute a hierarchy rather than a binary opposition (Gundel et al., 1993). In this dissertation, the most central information structural notion is be focus. The description of what is identified as focus here is given in (2). It is not meant as a delimitating definition of focus. Instead, it lists the relevant cases that will be understood as focus here.

[2] Of course, the existence of additional lexical tones in tone languages is recognised. However, these are often excluded from the analysis of intonation, which is commonly restricted to the post-lexical level per definition (see Ladd, 1996, p. 6, Gussenhoven, 2004, p. 12).

(2) In the answer of a question-answer pair, that part of the sentence is in focus which is

a. explicitly asked for in the question

b. or correcting it.

The notion of newness is often conflated with the idea of focus (see the discussion in Selkirk, 2007). On an intuitive level, focus is also often characterised as the 'important' part of the utterance or—maybe even more frequently—as 'new' (see the discussion in Krifka, 2007). This is captured in a more formal way by Lambrecht's (1994) definition of focus as "[t]he semantic component of a pragmatically structured proposition whereby the assertion differs from the presupposition" (1994, p. 213). In contrast, Rooth's (1985; 1992) widely-accepted semantic formalisation of focus as signalling that a set of alternatives is relevant in the interpretation of the utterance does without reference to newness. Krifka's (2007) definition, which is based on Rooth's (1985; 1992)'s theory of Alternative Semantics, is given in (3), with his first, more informal formulation in (3a) and his final definition in (3b).

(3) a. Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions (Krifka, 2007, p. 18).

b. A property F of an expression α is a Focus property iff F signals
(a) that alternatives of (parts of) the expression α or
(b) alternatives of the denotation of (parts of) α are relevant for the interpretation of α (Krifka, 2007, p. 19).

For the current purpose, a decision between Lambrecht's (1994) and Rooth's (1985; 1992) or Krifka's (2007)—and most other available approaches to focus is not necessary, but see Krifka (2007) for an explicit discussion of various alternatives. In the materials used for this thesis, utterances with different information structures were elicited with question-answer pairs (see sections 2.2.1 and 4.2.1). For these simplified mini-discourses, there is a good agreement on where the focused part is located, even when definitions of why this is to be understood as focus differ. Consider the example in (4). With Lambrecht, one could say that the presupposition here is that somebody wrote 'War and Peace' (x wrote 'War and Peace'), while the assertion of the answer is that that somebody was Tolstoy (x = Tolstoy). An account following Krifka would state that a set of alternatives are relevant to the interpretation of the answer in (4), which might include *Chekhov wrote 'War and Peace'*, *Dostoevsky wrote 'War and Peace'* and *Nabokov wrote 'War and Peace'*, but not *Tolstoy wrote 'Crime and Punishment'*. Out of the set of persons that could be x (the person who wrote 'War and Peace'), Tolstoy is identified as x. The outcome is the same with both approaches: The focus of the answer in (4) is on *Tolstoy*, as marked by square brackets.

(4) Q. Who wrote ‘War and Peace’?
A. [Tolstoy]$_F$ wrote ‘War and Peace’.

According to the description in (2), answers to questions like *What happened?* fulfil the criterion in (2a) and should be analysed as focused on a whole, as illustrated in (5) (but see for example Selkirk, 2007, for a different opinion). These cases, frequently called ‘all-new sentences’, will be referred to as ‘broad focus’ here to acknowledge the difference in scope compared to (4) (‘narrow focus’). At any rate, the materials used in the present context should in practice be fairly uncontroversial, as the broad focus examples were all-new sentences, whereas the non-focused parts of the narrow focus examples were always mentioned in the questions and thus also given.

(5) Q. Why do you look so tired?
A. [My children are ill]$_F$.

In addition to question-answer pairs like in (4) and (5), another kind of mini-dialogue was used in the materials of this thesis, as illustrated in (6). Whereas the answers in (4) and (5) provide information asked for in the question, the answer in (6) corrects the preceding question. The type of focus exemplified in (4) has been termed presentational or information focus, while (6) shows corrective focus, a subtype of contrastive focus.

(6) Q. Did Dostoevsky write ‘War and Peace’?
A. No, [Tolstoy]$_F$ wrote ‘War and Peace’.

The difference between these two types of focus is empirically relevant, since some languages have been reported to use different focus marking strategies for them (see e.g. Gussenhoven, 2004, pp. 180–183). The distinction has also been shown to be relevant for Finnish from a syntactic point of view. Vilkuna (1989, 1995) argues that Finnish sentences can be divided into K-position, T-position and V-field, as illustrated in Table 1.1 with an example by Jokinen (2005). Thereby, the V-field usually starts with the finite verb and the other two positions are (optionally) filled by the constituents preceding it. With some reservations, Vilkuna (1989, 1995) generalises that the sentence topic is realised in the T-position and the K-position is the default position for contrastive elements (topics or foci), whereas non-contrastive foci or ‘main news’ are normally located towards the end of the V-field (also see Vallduví and Vilkuna, 1998, who split up the notion of focus into rheme and kontrast).

1.3 Finnish segmental inventory

The native Finnish phoneme inventory consists of eight vowels and thirteen consonants. Almost all can appear in phonemically long or short versions,

Table 1.1: An example for the connection between word order and information structure in Finnish, based on Jokinen (2005, p. 229). The constituents are *karhu* ‘bear (nominative)’, *pyydysti* ‘caught’ and *kalan* ‘fish (accusative)’.

K-position	T-position	V-field	English equivalent
	Karhu	*pyydysti kalan*	‘The bear caught the fish.’
	Kalan	*pyydysti karhu*	‘The fish was caught by the bear.’
Kalan	*karhu*	*pyydysti*	‘It is the fish that the bear caught.’
Karhu	*kalan*	*pyydysti*	‘It is the bear that caught the fish.’
Pyydysti	*karhu*	*kalan*	‘The bear did catch the fish.’

whereby long phonemes are represented as double letters orthographically (a short account of lexical quantity in Finnish is given in section 1.5 below). In the following, Finnish expressions will usually be rendered using the standard orthography, which is near-phonemic.

The native consonants are traditionally given as [p t k d s h v j l r m n ŋ], written *p t k d s h v j l r m n ng/nk* (see Iivonen, 1998, p. 312, Karlsson, 2005, p. 66, but also see Karvonen, 2005, pp. 17–20 and Suomi et al., 2008, pp. 23–38). Following Suomi et al. (2008), the segment traditionally described as [v] is better classified as [ʋ]. Note also that [d] and [ŋ] are only found word-internally in native words, and [ŋ] appears either in combination with the velar plosive [k], spelled *nk*, or as a geminate, spelled *ng*; compare the nominative *renki* [reŋki] ‘farmhand’ vs. its genitive *rengin* [reŋːin] Additionally, the consonants [b], [f], [g] and [ʃ] can appear in recent loanwords and colloquial language, e.g. *banaani* [bɑnɑːni] ‘banana’, *filmi* [filmi] ‘movie’, *gluteeni* [gluteːni] ‘gluten’ and *šakki* [ʃɑkːi] ‘chess’ (see Karvonen, 2005, pp. 17–20).

The eight native vowel phonemes, written *i e ä y ö u o a*, are usually given as [i e æ y œ u o a] (see Karlsson 2005, p. 66, Harrikari 2000, p. 2), although for example Iivonen (1998, p. 312) and Karvonen (2005, p. 17) use [ø] instead of [œ]. Based on an acoustic investigation of F1/F2 formant values, Iivonen and Harnud (2005, p. 65) conclude that it would be more appropriate to render the short vowel phonemes as [i ɛ̝ a y œ̝ ʊ ɔ̝ ɑ] and the long ones as [iː e̞ː aː yː ø̞ uː o̞ː ɑː]. They do, however, note that this notation would slightly exaggerate the differences between the mid vowel versions (see ibid.; also cf. the discussion in Suomi et al., 2008, p. 20-21). When necessary, the

Table 1.2: Finnish consonant phonemes according to Suomi et al. (2008). Brackets mark segments with restricted distribution.

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Palatal	Velar	Glottal
Plosive	p (b)		t̪ (d)				k (g)	
Nasal	m		n				(ŋ)	
Fricative		(f)		s	(ʃ)			h
Approximant		ʋ				j		
Lateral approximant			l					

Table 1.3: Finnish vowel phonemes according to Suomi et al. (2008). Vowel symbols to the left of bullet points represent unrounded vowels, symbols to the right of bullet points represent rounded vowels.

	Front	Central	Back
Close	i • y		• u
Close-mid	e • ø		• o
Open-mid			
		æ	
Open			ɑ •

conventions used by Suomi et al. (2008) will be used here, transcribing the consonants as [p b t̪ d k g f s ʃ h m n ŋ r l ʋ j] and vowels as [i e æ y ø u o ɑ]. A more detailed discussion of the segmental inventory of Finnish can for example be found in Karlsson (1983, pp. 52–70), Iivonen (1998, p. 312), Karvonen (2005, pp. 17–20) and (Suomi et al., 2008, pp. 20–38).

1.4 Finnish prosodic hierarchy

This section summarises the prosodic hierarchy assumed in the present thesis on the basis of the research presented here and in previous literature. The current approach assumes that at least the units mora, syllable, prosodic word, prosodic phrase (p-phrase), intonational phrase (i-phrase), and possibly the foot, are relevant for the description of Finnish. These domains will be described in bottom-up order, starting with mora and syllable (section 1.4.1), then proceeding to foot and prosodic word (section 1.4.2), and finally discussing phrase-level prosody (section 1.4.3).

The present description concentrates on standard Finnish, the variety spoken in schools, radio and TV, which is close to the written language. Standard spoken Finnish will also generally be used by participants of linguistic studies when speaking sentences in a recording studio, although the dialectal background of the speaker will still have an influence (Karlsson, 1999; Ylitalo, 2009; Suomi, 2009). Therefore, participants from Helsinki and its surroundings were recruited for the three main experiments on which the thesis is based. Prosodic research comparing different Finnish dialects, which are generally mutually intelligible without problems, can be found in Wiik (1975, 1985, 1988), Iivonen and Yli-Luukko (1991) and Ylitalo (2009).

1.4.1 Mora and syllable

Syllables in basic native Finnish can be mono- bi- or trimoraic, as the initial syllables of the examples in Table 1.4 illustrate (see also Karlsson, 1983, p. 134–138, whose description is largely based on Wiik, 1977). Thereby, syllable onsets are non-moraic, whereas short vowels and coda consonants carry one mora each. Long vowels and diphthongs are associated with two moras. Geminate consonants are always ambisyllabic and thus contribute one mora to the weight of the first syllable. Note that here and in the following, long vowels and diphthongs are notated as VV, while geminates and consonant clusters are rendered CC in accordance with the Finnish phonological tradition, although an autosegmental framework is assumed here (see Harrikari, 2000, for a comparison of phonological representations of segment length in Finnish).

Finnish syllables can consist of onset, nucleus and coda, with the nucleus being the only obligatory part. Since the nucleus position can be filled by a short vowel, the minimal syllable can thus be notated as V (see ***i**.lo*

Table 1.4: Basic Finnish syllable types.

Weight	Structure	Example
Monomoraic	(C)V	*sä.e* 'verse'
Bimoraic	(C)VV	*lii.te* 'attachment'
		lai.te 'device'
	(C)VC	*ras.kas* 'heavy'
		hat.tu 'hat'
Trimoraic	(C)VVC	*haus.ka* 'pleasant'
		sääs.tä 'to save'
	(C)VCC	*pals.ta* 'column'
		särk.kä '(sand)bank'

'joy', *ko.**e*** 'test'). While onsets maximally consist of a single consonant in the core native vocabulary, codas can be complex, containing maximally two segments (on possible codas and word-internal consonant clusters, see Karlsson, 1983, pp. 107–127). Complex codas can however not co-occur with nuclei consisting of a long vowel or diphthong, i.e. fourmoraic syllables do not occur in basic native words (see Karlsson, 1983, p. 134).

In loan words, names and slang, syllable structure can be more complex and VVCC rhymes are possible (e.g. ***juots**.ka* from *juo.ta.va* 'something to drink'). Also complex onsets occur in names (e.g. ***Klauk**.ka.la*, a village in Southern Finland), colloquial spoken language and slang (e.g. ***sta**.din **frii**.du* 'native Helsinki woman') and loan words (e.g. ***pro**.fes.so.ri* 'professor', on the phonology of loanwords, also see Karvonen, 2005, pp. 25–26). However, Karlsson states that the two phenomena cannot co-occur as *CCVVCC, not even in marginal cases, as he exemplifies with **klauts.ka*, which is not a possible nickname for Klaukkala (Karlsson, 1983, p. 134).

Karlsson (1983, p. 134) relates that monomoraic syllables are traditionally called light / short, while all syllables with more moras are classified as heavy / long. In contrast to this, Harrikari (2000, 2003) has argued that CVV syllables are heavier than CVC syllables, although she does not formalise this hypothesis in terms of moraic structure (Karvonen, 2005, also supports this position, likewise without reference to mora assignment). This thesis assumes a tertiary distinction between light / short monomoraic syllables, heavy / long bimoraic syllables and superheavy / overlong trimoraic syllables. Although this assessment should be verified by further research, the timing patterns reported in Chapter 4 suggest that analysing moras as timing units, a three-way distinction emerges (evidence for the mora as a timing unit in Finnish also comes from O'Dell et al., 2007).

1.4.2 Foot and prosodic word

This section discusses the foot and prosodic word (PW), starting with the latter unit. Prosodic words constitute the domain of several phonological processes in Finnish. First, vowel harmony and other phonotactic restrictions apply to the prosodic word, as mentioned by Karlsson (1977).[3] Finnish vowels can be divided into three groups: front harmonic vowels *y ö ä* [y ø æ], back harmonic vowels *u o a* [u o ɑ], and harmonically neutral vowels *i e* [i e]. Front and back harmonic words cannot co-occur in the same word, while harmonically neutral vowels may appear together with vowels from either of the two other groups. Thus, ***talo*** [t̪ɑlo] 'house', ***tila*** [t̪ilɑ] 'condition', ***työ*** [t̪yø] 'work', ***tyyli*** [t̪yːli] 'style', ***tuo*** [t̪uo] 'that' and ***tuuli*** [t̪uːli] 'wind' are existing words, but ****talö*** [t̪ɑlø] and ****tyo*** [t̪yo] are not. Also, suffixes either contain only harmonically neutral vowels or exist in a back harmonic and a front harmonic variant (roots containing only the neutral vowels *i* and *e* combine with front harmonic suffix versions); as an example, compare the adessive forms *tal**olla*** [t̪ɑlolːɑ] 'at the house' and *työ**llä*** [t̪yølːæ] 'at work' (for a more detailed discussion, see e.g. Suomi et al., 2008, pp. 51–54). Karlsson (1977) also mentions that the phonemes [ŋ] and [d], as well as about 60 consonant clusters appear only word-medially. For example, the sequence *lst* cannot be a complex onset or coda and cannot be dissolved across word boundaries into *-*l st*- or *-*ls t*-, but does appear in ***pals.ta*** 'column' (but cf. Suomi, 1985).

Second, morphophonological alternations like consonant gradation take place inside the word (see Karlsson, 1977, also see Karlsson, 1999, pp. 28–43). Consonant gradation affects the non-marginal plosives [p t̪ k] when they are in the onset position of the second syllable in a stem morpheme. This happens in one of two basic ways. First, the long versions [pː t̪ː kː] alternate with the short versions of the same plosives in certain grammatical conditions. For example, the long velar plosive in the nominative ***kukka*** 'flower' (the so-called 'strong grade') corresponds to a short velar plosive when the genitive / accusative marker *-n* is added, i.e. ***kukan*** ('weak grade'). Second, the short versions of the same consonants alternate either with other consonants or, in the case [k] only, with ∅ in the same grammatical conditions. For example, the genitives of the words ***tapa*** 'custom, habit', ***satu*** 'fairy tale' and ***sika*** 'pig' (strong grade), are ***tavan***, ***sadun*** and ***sian*** (weak grade), respectively. Consonant gradation also applies when [p t k] are part of a consonant cluster, i.e. preceded by another consonant in the coda of the first syllable; in this case, assimilation frequently applies, e.g. the genitive of ***kulta*** 'gold' is ***kullan*** (see Karlsson, 1999, pp. 28–43, for a full account of all the variants).

[3] Actually, Karlsson simply refers to 'the word', but at least his phonological observations will be interpreted as pertaining to the prosodic word here.

Third, the minimal word requirement demands that Finnish words—with the exception of a few function words—contain at least two vocalic moras (e.g. Hanson and Kiparsky, 1996; Suomi, 2004, 2005). At least in standard language, V syllables never form words of their own, and all three VC words (*en* 'I do not', *et* 'you (singular) do not' and *on* 'he / she / it is'), five CVC words (*hän* 'he / she', *kun* 'when', *jos* 'if', *nyt* 'now', *kas* 'look!, oh, my!') and seven CV words (*me* 'we', *te* 'you (plural or formal)', *he* 'they (animate)', *ne* 'they (inanimate)', *se* 'it', *ja* 'and' and *jo* 'already') are function words, i.e. belong to a closed class (Karlsson, 1983, p. 215). While there are only three VV words, two of which are content words (*ei* 'he / she / it do not, do not!', *yö* 'night', *ui* 'he / she / it swims, swim!'), there are about 50 CVV words, including about 40 verb and noun forms like *kuu* 'moon', *pää* 'head' and *myi* 'he / she sold' (Karlsson, 1983, pp. 215–216). CV.CV words, in turn, are numerous, with 756 uninflected noun forms alone (Karlsson, 2005, p. 69). Hanson and Kiparsky (1996) take the minimal word requirement as evidence for their suggestion of the moraic trochee as the foot involved in assigning stress in Finnish, while Suomi (2004, 2005) relates it to the need of space to associate accentual tones directly, while classifying Finnish as a syllabic trochee language (cf. the discussion of stress below).

Fourth, this thesis assumes that the prosodic word is the domain of syllabification in Finnish. Consider the contrast between the plural essive form *i.s**oi**.na* 'being great' and the compound *i.s**o**.**i**.sä* 'grandfather' illustrated in (7) and (8). Both include the sequence *oi* divided across a morpheme boundary. However, vowel harmony indicates that the stem and its suffixes in (7) form one prosodic word, as indicated by the bracketing, because the essive suffix appears in its back harmonic variant *-na* (as opposed to front harmonic *-nä*). The vowel sequence *oi* is consequently part of the same syllable and forms a diphthong. In contrast, the two stems in (8) each constitute a separate prosodic word under the present analysis, since the back harmonic vowel *o* and the front harmonic vowel *ä* cannot co-occur within one prosodic word. The intervening prosodic word boundary is taken to prevent *o* and *i* from being tautosyllabic.[4]

(7) (i.s**oi**.na)$_{\text{PW}}$

iso-	i-	na
great	PLURAL	ESSIVE

'(them) being great'

[4]Alternatively, it is possible to assume a recursive prosodic word structure as suggested by Ito and Mester (2009). Under such an analysis, (8) could be given as ((i.s**o**.)$_{\text{PW}}$(**i**.sä)$_{\text{PW}}$)$_{\text{PW}}$ and the minimal prosodic word could be assumed as the domain of vowel harmony and syllabification. However, recursivity in the prosodic structure will not be pursued further here due to the lack of suitable data, but see Karvonen (2005) for a suggestion using recursive prosodic words in the analysis of compound and simplex words.

(8) $(\text{i.s}\mathbf{o}.)_{\text{PW}}(\mathbf{i}.\text{sä})_{\text{PW}}$
iso- isä
great father
'grandfather'

Finally, virtually all accounts of Finnish phonology agree that the first syllables of (prosodic) words carry fixed primary stress (see e.g. Sadeniemi, 1965; Wiik, 1981; Karlsson, 1983; Välimaa-Blum, 1993; Iivonen, 1998). However, the present analysis does not put a strong emphasis on stress due to phonological as well as phonetic reasons. From a phonological point of view, the status of compounds and short function words would be unclear if the primary stress criterion were added to the definition of the prosodic word employed here. Although compounds do not function as a unified domain for vowel harmony, as illustrated above, they are usually described as forming one word with initial primary stress (see Sadeniemi, 1965; Wiik, 1981; Sulkala and Karjalainen, 1992; Karvonen, 2005, but also see Iivonen, 1998, p. 315). Similarly, short function words like *ja* 'and', *on* 'is', *se* 'it' or *kun* 'when' do not harmonise with the adjacent material in vowel backness, thus forming prosodic words of their own under the current analysis, but are frequently described as unstressed or merged with neighbouring words in continuous speech (see Karlsson, 1983, p. 150, Iivonen and Yli-Luukko, 1991, p. 37, Sulkala and Karjalainen, 1992, p. 381). The current analysis assumes that Finnish compound words are generally realised as p-phrases consisting of two or more PWs (see Chapter 5) and that also function words frequently form p-phrases with adjacent prosodic words. From a phonetic point of view, there is a lack of evidence for strong cues of primary stress, see the investigation in section 3.4 of this thesis.

That stress is less central to the prosodic system and / or less strongly marked phonetically in Finnish than for example in English is also asserted by Sadeniemi (1949), Niemi (1984), Iivonen (1998) and Ylitalo (2009). Note, however, that the overall prosodic analysis suggested in this dissertation does not depend on the issue of stress in any way, as will be discussed in Chapter 6.

The unit foot has been mentioned in connection with two phenomena, timing and stress. While primary stress seems to be relatively uncontroversial, albeit relatively unpronounced, agreement on secondary stress placement seems to be much more difficult to obtain, both in terms of native speaker intuition and with respect to its phonological analysis (Sadeniemi, 1949, 1965; Hanson and Kiparsky, 1996; Elenbaas and Kager, 1999; Karvonen, 2005; Karttunen, 2006; Anttila, 2010). Most of the these studies agree that odd-numbered syllables with the exception of the last one receive secondary stress, but syllable weight and morphology also play role. Anttila (2012) asserts that foot structure does not always have prosodic correlates and that it can be difficult to obtain reliable data on secondary stress, but does present a phenomenon of segmental variation (t-deletion) that can be accounted for with reference to foot structure. Also, Wiik and Lehiste (1968) and Lehtonen (1970) found that lexical quantity differences caused adjustments of segment durations within a unit that probably corresponds to the foot—although both studies refer to 'speech measures', so that the correspondence is not quite clear. Also note Suomi's (2007b) explicit statement that the relevant stretch is smaller than a foot and "does not constitute an independently motivated phonological / prosodic unit in longer than disyllabic words, except accidentally" (Suomi, 2007b, p. 54). With regard to the topics of this dissertation, there is no evidence for the relevance of the foot.

1.4.3 Phrase-level prosody

The analysis suggested in this thesis employs two prosodic domains above the prosodic word level: p-phrase and i-phrase. A p-phrase consists of at least one prosodic word, but often unites several prosodic words. Examples of constituents frequently forming p-phrases in the materials of this thesis include VPs (see Chapter 2), NPs (see Chapter 4) and compounds (see Chapter 5). P-phrases are most prominently marked by two tones, $\mathrm{H_p}$ and $\mathrm{L_p}$. $\mathrm{H_p}$ is usually realised early in the p-phrase, while $\mathrm{L_p}$ normally appears towards its ends, but some variation in timing is found.

The i-phrase is the highest prosodic domain discussed in this analysis. While higher-ranking units like the utterance are doubtlessly relevant in Finnish prosody as well, the present data corpus consists of relatively short sentences and does therefore not lend itself to a detailed investigation of these units. The data presented in Chapter 2 indicate that short simple sentences per default form a single i-phrase, but that this default is affected by other factors such as information structure. The findings presented there suggest that the right edge of a focused constituent is aligned with the right edge of an i-phrase. They further indicate that these edges are signalled by pauses, final lengthening and non-modal voice quality. Additionally, Chapter 3 suggests that the i-phrase is optionally marked by either a low or a high final boundary tone ($\mathrm{L_i}$ and $\mathrm{H_i}$, respectively). In contrast to many other languages, the choice of a final high or low tone does however not indicate a difference between questions and statements. Finnish questions do generally not end with rising pitch, and it has been debated whether they are prosodically different from statements in a principled way at all, see Hirvonen (1970); Iivonen (1978); Anttila (2008). The intonation of questions will therefore not be discussed further in this thesis.

The analysis suggested here differs from existing accounts characterising Finnish intonation. While the relevance of the i-phrase—under different names—is implied or stated in most prosodic descriptions (explicitly by Välimaa-Blum, 1988, 1993), the main difference is the assumption of a p-phrase marked by $\mathrm{H_pL_p}$ tones. In general, agreement on prosodic domains larger than a word and smaller than an i-phrase is more difficult to obtain (for an overview of suggestions regarding Finnish, see Iivonen et al., 1987, pp. 226–228). Crucially, pitch movements that are analysed as realisations of H_pL_p tones here have so far been described in terms of pitch accents realised on stressed syllables (e.g. Iivonen, 1998). A common striking feature is that these accents are usually described as uniformly rising-falling. For example, Iivonen (1998) introduces a distinction between three different types of accents—accent for rheme, accent for contrast and accent for emphasis—but does not mention any difference in their f_0 contours (but see Iivonen, 1993, 1998, p. 320, for a description of marginally occurring different shapes with expressive or stylistic functions).

While most of the literature on Finnish intonation refrains from explicitly espousing a particular theoretical model, two descriptions of Finnish intonation have so far referred to a bitonal framework. First, Välimaa-Blum's (1988) AM-account uses two boundary tones that are associated with the end of the intonational phrase: L% to model the usual final pitch fall and H% to model the continuation contour, which is not rising, but characterised by the absence of the fall. She suggests that these boundary tones and one pitch accent, L+H*, are sufficient to model the "basic intonational structure" (Välimaa-Blum, 1988, p. 114) of Finnish, including various clause types in different word orders, and expects that no radical adjustments would be necessary to extend the analysis to extragrammatical uses of intonation (see Välimaa-Blum, 1988, p. 102). A second pitch accent, L*+H, is added in Välimaa-Blum (1993) to explain certain—obviously rather marginal—contour patterns. While some examples of its occurrence are given, Välimaa-Blum does not discuss whether or how this second accent is functionally different from the usual L+H*. She does, however, explicitly state that focus results in an increased pitch range, while the accent used is identical to those used in sentences without (narrow) focus, i.e. also L+H* (Välimaa-Blum, 1993, pp. 89–90).

Second, Suomi et al. (2008) have argued that Finnish accents should be described as LHL without a starred tone, since they found all three tones to have a stable segmental anchor point (see Suomi et al., 2008, p. 82, 122, Suomi, 2007a, 2009). However, Ylitalo (2009, p. 135) suggests that Suomi 2009's results could be formalised with the notation L+H*L. Suomi et al. (2008) explicitly disagree with Välimaa-Blum (1993) and state that "the Finnish accentual tune, at least in Northern Finnish, is usually a rise-fall, with the fall clearly being an essential part of the tune" (Suomi et al., 2008, p. 122). A central claim of their work, first fully formulated in Suomi (2005), is that Finnish accents are tonally uniform across word structures and that segment durations are adjusted "in order to enable uniform tonal realization of accent" (Suomi, 2005, p. 298). The uniform tonal movement is characterised in bitonal terms as an L tone anchored to the word-initial consonant, an H tone anchored to the end of the word's first mora and an L tone occurring close to the end of the second mora (Suomi, 2007a, p. 112) or around the middle of the third syllable, with its precise anchoring being a bit unclear (Suomi et al., 2008, p. 80). Other articles use the formulation of a rise on the first mora and a fall on the second mora and following segments (Suomi et al., 2003; Suomi, 2005, 2007b, 2009).[5]

1.5 Prosodic marking of lexical quantity in Finnish

There are two levels of quantity for all Finnish vowels and most consonant phonemes in almost all positions in a word, e.g. *muta* 'mud', *mutta* 'but', *muuttaa* 'to change', *ei muuta* 'does not change', *mutaa* 'mud (partitive)', *odottaa* 'to wait', *odota* 'wait!'. However, long consonants are always ambisyllabic, so that *kor.pi* 'wilderness' and *korp.pi* 'raven' are distinct words in Finnish, but **korr.pi*, **kor.ppi* and **korr.ppi* are not possible.[6] The current dissertation will use the terms 'short' and 'long' or 'quantity 1 (Q1)' and 'quantity 2 (Q2)' for the two levels.

There are two prosodic cues for the quantity distinction: duration and pitch. Lehtonen's (1970) extensive study showed that while long segments are generally speaking about twice as long as short ones, segmental durations are crucially adjusted within the foot or *(puhe)tahti* '(speech) measure' (see also Wiik and Lehiste, 1968, but also cf.the discussion in section 1.4.2 above) The most conspicuous phenomenon is the so-called half-long vowel, which occurs in the second syllable of (9a). The second vowel in a CV.CV structure will be lengthened subphonemically, but noticeably—a phenomenon of which speakers are all the more aware since it shows considerable dialectal variation (Wiik and Lehiste, 1968; Wiik, 1975, 1985; Ylitalo, 2009). The half-long vowel in (9a) thus differs from the other vowels in (9), which are either long or short. Another durational adjustment not explicitly marked in (9) is that the [t̪] in (9a) is shorter than in (9b) (see Lehtonen 1970, pp. 122–124, also see Suomi 2009).

(9) a. *sata* [sɑt̪ɑˑ] 'hundred (nominative)'
    b. *sataa* [sɑt̪ɑː] 'hundred (partitive)'

[5]In recent publications, Suomi and his colleagues describe their analyses as pertaining to Northern Finnish and possibly not completely generalisable to other varieties (Suomi, 2007a, 2009). However, their findings will be treated as relevant to this thesis, since there is strong indication that the central claim of uniformity does apply to other dialect regions. While Ylitalo (2009) finds differences for South-Western Turku speakers, she confirms uniformity not only for the Northern variety as spoken in Oulu, but also for the speakers from Tampere, which belongs to the Central or Häme dialect region. Already earlier, Wiik (1988) observed uniform peak timing as seen from the beginning of the word's first vowel for his Häme speaker, a woman from Kotka, in a small corpus of sentences with different word structures. These findings indicate that the uniformity hypothesis should also be applicable to the current data, as Helsinki is also part of the Häme dialect region.

[6]Recall that the representation of long phonemes with double graphemes is chosen purely for typographical convenience here.

c. *ei saata* [sɑːt̪ɑ] *olla* 'might not be'

d. *saattaa* [sɑːt̪ːɑː] *olla* 'might be'

In addition to duration, also pitch is relevant in distinguishing the two levels of quantity—at least in the first syllable (Malmberg, 1949; Lehtonen, 1970; O'Dell, 2003; Vainio et al., 2010). The tonal cues are also exploited in language processing, as Järvikivi et al. (2010) have shown, so that an ambiguous vowel duration will more likely be perceived as either long or short depending on the tonal contour it carries. Vainio et al. (2010) formalise the difference with reference to Xu's (1999; 2005) target approximation model and state that the first syllable has a high target in quantity 1, but a falling target in quantity 2. The current thesis employs another way of characterising the pitch cues to the quantity difference. This approach, advocated by O'Dell (2003) and Suomi (2007a, 2009), can be termed the time line model. It assumes that the phonological specification of the pitch contour is the same for words with different syllable structures, but realised across segments with different durations. The model explains the perceptual relevance of pitch cues to quantity as a time line. Thereby, the uniform tonal movements provide a help in identifying durations as phonologically short or long, in that listeners can determine with respect to the fixed tonal shape how much relative time has passed, irrespective of speech tempo (Suomi, 2007a, 2009). For example, when a vowel in the first syllable of a word is realised with falling pitch (analysed here as a realisation of $\mathrm{H_pL_p}$ tones), the listener can infer that this vowel has a long duration, since it provided enough time for both targets to be realised. The vowel is thus interpreted as bimoraic and long. In contrast, simple high pitch on the same vowel (within the limits of ambiguous vowel duration) can indicate that this vowel is short, since it does not provide enough time and moraic space for the realisation of the $\mathrm{L_p}$. Thus, O'Dell has suggested that "pitch movement helps to provide the listener with an indication of the speaker's "time line" (i.e. helps synchronize listener and speaker) for the purpose of making quantity judgements" (2003, pp. 77–79).

1.6 Prosodic marking of information structure in Finnish

The pitch contour of neutral declarative Finnish sentences is generally characterised as a series of rising-falling accents which decrease in f_0 height, frequently ending in creaky voice (Iivonen et al., 1987; Iivonen, 1998). In these contours, basically all content words are said to bear an accent (Sulkala and Karjalainen 1992, p. 382, Iivonen 1998, p. 317), although finite verbs have been described as less prominent, sometimes or always unaccented (Mixdorff et al., 2002; Välimaa-Blum, 1988, 1993, for a somewhat inconclusive experimental investigation see Arnhold et al., 2010). Under the analysis suggested here, these neutral pitch contours are analysed as realisations of $\mathrm{H_pL_p}$ tones showing regular downstep, with the verb not consistently forming a p-phrase of its own, but usually uniting with neighbouring constituents.

That the neutral declarative pattern is altered and intonation is used to express information structure has been attested in several studies (Välimaa-Blum, 1988, 1993; Mixdorff et al., 2002; Vainio and Järvikivi, 2006, 2007): A word in narrow focus is realised with a greater pitch range. Thereby, Vainio and Järvikivi (2006, 2007) found that the rise to the peak is important in pre-final position, whereas utterance-finally, the fall is more important. Mixdorff et al. (2002) report that pre- and post-focal accents were never completely deleted in their study, but e.g. Suomi et al. (2003) and Suomi (2005) describe unaccented words in the context of contrastive focus (also compare Välimaa-Blum, 1988, pp. 110–114). The additional use of duration (Mixdorff et al., 2002; Suomi, 2007b) and intensity (Vainio and Järvikivi, 2007) for marking focus have been reported, although the effect of intensity is weaker than that of pitch in perception (Vainio and Järvikivi, 2006) and according to Suomi et al. (2008) accentual lengthening is reserved for contrastive accents. Also, Vainio and Järvikivi (2006) have shown that prosody and syntax interact in marking information structure. Although differing with respect to some details, the data presented in Chapters 2, 3 and 4 overall confirm previous findings regarding the effect of information structure on pitch range, duration and intensity. Additionally, they identify information structure effects for the use of voice quality and pauses.

Crucially, previous instrumental studies of prosodic variation introduced by information structure or prominence differences do not propose that Finnish would utilise different types of accent in the sense of different combinations of tones and / or different associations between tones and segmental structure. On the contrary, authors either discuss prosodic effects without mentioning changes in accent type as defined by f_0 shape and tonal timing (Iivonen, 1998; Vainio et al., 2003; Vainio and Järvikivi, 2006, 2007; Suomi et al., 2008) or state explicitly that narrow focus constituents are realised with the same accent as words in other information structural conditions (Välimaa-Blum, 1993; Mixdorff et al., 2002). The present thesis likewise does not assume different pitch accents for sentences with different information structures, but ascribes prosodic differences to variation in phrasing.

1.7 Thesis overview

The main hypothesis of this thesis is that Finnish is a phrase language in the sense of Féry (2010). The defining characteristic of these languages is that their intonation is determined by a few, relatively invariant phrase-level tones instead of being defined by a set of lexical tones or by pragmatically governed choices from an inventory of pitch accents. Thus, phenomena that have traditionally been described as related to stress and / or accent in Finnish phonology are analysed here as shaped by variation in prosodic phrasing and the resulting distribution of phrase tones. In particular, what has been described as a uniform accent—whether L+H* (Välimaa-Blum, 1993), rise-fall (Iivonen, 1998) or LHL (Suomi et al., 2008)—is analysed as a realisation of tones associated with the p-phrase (H_pL_p), while effects of information structural variation are accounted for primarily on the i-phrase level.

The proposed analysis is mainly based on data from three experiments designed to address the three main research questions given in (10).

(10) a. Is Finnish phrase-level intonation indeed largely uniform or are there meaningful phonological contrasts that warrant the assumption of contrasting accents?

b. What is the shape and tonal specification of the pitch contour(s)?

c. What is the position of Finnish in a prosodic typology?

Two production experiments were designed primarily to address the first research question. For both, the underlying rationale was to systematically vary parameters that were expected to induce the realisation of different types of accent, if this option existed in Finnish intonation. The first study, Production Experiment I described in Chapter 2, manipulated (first syllable vowel) quantity, position in the sentence and information structure. As summarised in section 1.5, one cue to the lexical quantity distinction is pitch. Since the difference has been described as involving different underlying pitch targets (Vainio et al., 2010), it was included in the materials to eximine its effects and possible interaction with other factors. Position and information structural variation were investigated because they are generally seen as central in the selection from an inventory of pitch accents, i.e. most accounts of intonation languages describe an opposition between nuclear and pre-nuclear accent and / or pitch accents associated with topics vs. focused constituents (e.g. Gussenhoven, 1983 for English; Féry, 1993 for German; Gussenhoven, 2005 for Dutch). While Production Experiment I identified prosodic effects of all three manipulated factors, there was no clear evidence for contrasting accents. The chapter outlines an analysis in terms of phrase tones, but an alternative analysis in terms of pitch accents, the nuclear accent hypothesis, is formulated. Chapter 3 tests this hypothesis in more detail and refutes it based on a more thorough analysis of peak timing in the materials from Chapter 2, a pilot study on the perception of peak timing and an analysis of intensity as a marker of stressed syllables.

Following the same rationale as the first experiment, the second production study addressed the question of quantity and syllable structure in more detail (Chapter 4). While variation in information structure and sentence position of the target items were reduced, quantity was varied systematically not only in the first but also in the second syllable. In addition to open syllables with short or long vowels, different types of closed syllables occurred. Also, words were either di- or trisyllabic. Table 1.5 provides an overview of similarities and differences in the design of the two studies. Crucially, both experiments did not provide grounds for establishing an accent inventory. Instead, they confirmed previous descriptions of uniformity.

Table 1.5: Comparison of the design for the two production experiments: Conditions of the target words.

Factor	*Levels in Production Experiment I*	*Levels in Production Experiment II*
Position	(1) Sentence-initial (2) Sentence-medial (3) Sentence-final	(1) Pre-final
Information Structure	(1) Broad focus (2) Narrow focus (3a) Unfocused given, pre-focal (3b) Unfocused given, post-focal	(1) Broad focus (2) Narrow focus
Number of syllables	(1) Two	(1) Two (2) Three
First syllable structure (quantity)	(1) CV (short) (2) CVV (long)	(1) CV (short) (2) CVV(long) (3) CVC_g (long, geminate consonant) (4) CVC (long, non-geminate consonant) (5) CVVC (overlong)
Second syllable structure (quantity)	(1) CV (short) (2) CVV (long)	(1) CV (short) (2) CVV(long)

While the two production experiments mainly targeted the research question in (10a), their data also served to form hypotheses regarding (10b). To test these hypotheses regarding shape and phonological specification of pitch contours was the main purpose of the processing study presented in Chapter 5. It investigated which tonal movements were necessary for native listeners to perceive a word as having a tonal contour of its own. Results support an analysis using H_pL_p phrase tones.

The last research question will mainly be addressed in Chapter 6 on the basis of the findings of all three studies and some additional material. Overall, the results indicated Finnish to be quite different from typical intonation languages. The chapter provides a discussion of these findings and concludes with an overview of Finnish prosody under the suggested phrase language analysis.

Chapter 2

Production Experiment I: The influence of information structure, quantity and position

2.1 Introduction

This chapter presents a production experiment that was designed as a first step in tackling the research questions outlined in the previous chapter by dividing them into sub-questions and applying them to a relevant data set. Specifically, the study investigates whether there is evidence for a phonological contrast between tonal contours of stressed syllables and thus for different types of accent in Finnish (question (10a), reformulated as **Qu1** below). A follow-up question of how to model the emerging patterns appears in **Qu2**. The chapter also addresses the question of tonal specification formulated in (10b), repeated as **Qu3**.

Qu1 Can the data be accounted for by just one type of tonal specification on the phonological level or do they suggest a contrast between two or more accents?

Qu2 Is an analysis in terms of phrase tones or accents preferable for this data?

Qu3 What is the shape and tonal specification of the contour(s)?

The key idea of the experiment was to use manipulations that would probably result in the emergence of different pitch accents for a language with an accent inventory to choose from. Therefore, the design combined and systematically co-varied within one single study what can most plausibly be assumed to be important factors affecting tonal realisation in Finnish. The aim was to elicit different types of accents, if this possibility should exist in Finnish, or otherwise to study the variation in the realisation of the single tonal pattern. To be clear, these manipulations were not expected to elicit all possible pitch patterns. The goal was not to delineate a complete pitch accent inventory, it was to see if systematic variation indicating the existence of such an inventory could be elicited. Three factors were chosen here on the basis of previous literature on Finnish and other languages: information structure, position / grammatical function and quantity.

Almost all works dealing with intonation give examples of how it is employed to mark information structure. Also for Finnish, an influence of information structure, more precisely focus, has been illustrated in previous research (see Mixdorff et al., 2002; Vainio et al., 2003; Vainio and Järvikivi, 2006, 2007). Of particular importance in the present context is Vainio and Järvikivi's (2006; 2007) finding that different parts of the rise-fall contour are crucial for signalling narrow focus or prominence in different positions in the sentence. In their production and perception data, the rise was more important for marking narrow focus in pre-final position, while the fall was more important for sentence-final narrow focus. If this pattern were reproduced, it could provide evidence for a distinction between falling accents and rising accents.

A matter tightly connected with information structure is word order, which is in turn interwoven with grammatical function. As outlined in section 1.2.2, variations in word order imply differences in information structure in Finnish, with the finite verb often serving as an anchor point. The finite verbs themselves have traditionally been described as weakly accented or accentless unless narrowly focused (see Välimaa-Blum, 1993). This has to a certain extent been experimentally verified in a study by Arnhold et al. (2010), which also provided indication that the effect is indeed due to grammatical function and not attributable to linear position. However, the present experiment kept word order constant, so that grammatical function and position were investigated in connection with each other.

As described in section 1.5, lexical quantity is not only marked by duration, but by pitch cues, as demonstrated in a number of studies (Lehtonen, 1970; O'Dell, 2003; Vainio et al., 2006; Järvikivi et al., 2007; Vainio et al., 2010; Järvikivi et al., 2010). How to account for these pitch cues has been somewhat debated: They have been modelled as differences in underlying tonal specification (i.e. as a static high tone for quantity 1 vs. a falling tone in quantity 2 in Vainio et al., 2010), but also as the same tonal targets realised on durationally different material (O'Dell, 2003; Suomi et al., 2008, cf. section 1.5). The present experiment might be able to shed some light on this debate by taking additional factors into consideration. At any rate, it seems indicated to take quantity into account as a factor with an absolute or relative influence on tonal realisations.

2.2 Methods

2.2.1 Materials

The material was based on eight short sentences in unmarked SVO word order (see Table 2.1) embedded in different information structural contexts. All words—subject, verb and object of each sentence—were analysed and were thus target words. Each word in the materials consisted of two open syllables with onset consonants.

Vowel quantity was systematically manipulated and four types of quantity combinations occurred: quantity 1 in both first and second syllable (Q1Q1), quantity 1 in the first and quantity 2 in the second syllable (Q1Q2), quantity 2 in the first and quantity 1 in the second syllable (Q2Q1) and quantity 2 in both syllables (Q2Q2). Due to morphological requirements, the type Q2Q1 appeared only for subjects and Q1Q2 only for objects, while Q1Q1 occurred for subjects and verbs and Q2Q2 for verbs and objects. The subjects, first names in nominative case, always had a short second syllable vowel (see Karlsson, 2005, pp. 67–69, on the extreme rarity of long second syllables in Finnish disyllabic nominative singular nouns). The objects ended with a long syllable due to the partitive marker -*ä* / *a* [æ / ɑ] elongating the second syllable vowel, e.g. *la.va* [lɑ.ʋɑ] 'platform (nominative)' becoming *la.vaa* [lɑ.ʋɑː] 'platform (partitive)'.[1] The second position (verbs) was thus the only one in which vowel quantity was manipulated in both syllables simultaneously, while for subjects and objects, only first syllable vowel quantity varied. Of the sentences with long initial syllables (1–4 in Table 2.1), two contained only long monophthongs and two contained only diphthongs.[2]

From the semantic point of view, the sentences were designed so that the verb and the object would not be too expectable from each other; for example combinations like 'knit' and 'sock' or 'read' and 'book' were avoided. The lexical material was additionally chosen to be as sonorant as possible, although this led to the inclusion of two segments that can be difficult to segment: The consonant [j] occurred three times in word-initial and once in word-medial position, while [ʋ] was used twice word-internally and once as the initial consonant. Still, it was necessary to include verbs containing the obstruents [h], [t̪] and [k]. One of the verbs with long vowels and all four

[1] This description accounts for the object items used here, but note that Finnish partitive formation can involve stem alternation and / or the alternative partitive morpheme -*tä* / *ta* (compare Karlsson, 1999, pp. 44–54).

[2] In fact, the vowel combinations *ea*, *yä* and *ua* [eɑ yæ uɑ] that occur in the second syllable of the objects of sentences 3, 4, 6 and 7 are not usually viewed as diphthongs, but as belonging to two different syllables (Wiik, 1977; Karlsson, 1983), so that the relevant items should be syllabified as *loi.me.a*, *lyi.jy.ä* and *le.lu.a*. However, these words will still be treated as disyllabic here, since they did not significantly differ from the true diphthongs in terms of duration and it was as difficult to divide these vowel sequences into two parts as it was for the true diphthongs; see Appendix B for a discussion.

Table 2.1: Lexical material used in Production Experiment I.

	Subject	Verb	Object	Translation
1	Niilo	maalaa	liinaa	"Niilo (m) paints (a) cloth."
2	Moona	liimaa	naavaa	"Moona (f) glues lichen."
3	Maini	neuloi	loimea	"Maini (f) knitted a blanket."
4	Väinö	jauhoi	lyijyä	"Väinö (m) ground lead."
5	Jani	töni	lavaa	"Jani (m) pushed a platform."
6	Jimi	luki	menua	"Jimi (m) read the menu."
7	Manu	hali	lelua	"Manu (m) hugged a toy."
8	Jali	nyki	lanaa	"Jali (m) tugged a levelling drag."

verbs with short vowels had an obstruent as the onset of either the first or the second syllable.[3]

Each of the eight sentences in Table 2.1 was recorded with seven different information structures: with broad focus (all-new sentence), with narrow information focus on the subject, verb and object, respectively, and with narrow corrective focus on each of the three constituents, respectively. The information structures were induced by questions, e.g. for the sentence *Maini neuloi loimea* ('Maini knitted a blanket') the broad focus version was an answer to the question *Miten iltapäivä meni?* ('How did the afternoon go?'), corrective focus on the verb was elicited with the question *Silittikö Maini loimea?* ('Did Maini iron a blanket?') and information focus on the object with the question *Mitä Maini neuloi?* ('What did Maini knit?').

It should be pointed out that whenever one word was focused, the other two were not only part of the background, but also mentioned in the context question ('given'). Thus, words in each position were uttered in three basic focus conditions (disregarding the difference between information focus and corrective focus):

1. Broad focus and contextually new
2. Narrow focus and contextually new
3. Background and given (in the following: 'unfocused')

For example, when the sentence *Manu hali lelua* ('Manu hugged a toy') was uttered as an answer to the question *Mitä sitten tapahtui?* ('What happened then?'), all three words were in the broad focus condition. When the same sentence answered the question *Erkkikö hali lelua?* ('Did Erkki hug a toy?'), the subject *Manu* was in the narrow corrective focus condition, while the verb *hali* and the object *lelua* were both unfocused. Since the

[3]Indeed, the only two completely sonorant verbs with the desired syllable structure that I was able to find were *meni* 'he / she / it went' and *muni* 'he / she / it laid an egg'.

recorded sentences contained three words which were focused in turn, each word occurred twice as often in the condition unfocused as in narrow focus. Also, two different types of narrow focus—corrective focus and information focus—were elicitated, while there was only one kind of broad focus, an all-new sentence. Therefore, the ratio between the occurrence of the basic focus conditions broad focus, narrow focus and unfocused for each word was 1:3:6.

Note also that the position of a word in the sentence was correlated with its grammatical function, in that the subject always occurred in the sentence-initial position, followed by the verb in the second and the object in the final position. Thus, possible effects of the two factors cannot be distinguished. It would have been possible to avoid this ambiguity by varying word order; however, there is a strong connection between word order and information structure and the perception of prominence in Finnish (see Vilkuna, 1989, 1995; Vainio and Järvikivi, 2006, see also section 1.2.2): While the unmarked word order SVO is natural with all the information structures induced in the present experiment, changing the positions of the words, especially the verbs, would have resulted in unnatural combinations or, leaving these out, an unbalanced experimental design. More importantly, varying word order would have added an additional factor to an already complex experiment.

In sum, the combination of the eight sentences with different lexical items (four with long and four with short first syllable vowels) with the seven information structures led to a number of 56 sentences per speaker.

2.2.2 Participants

This chapter is based on data from 17 native speakers of Finnish (10 female, 7 male) from Helsinki or the surrounding area. They were 20 to 35 years old (mean 25.41). All were or had recently been students of various subjects at the University of Helsinki; none of them was familiar with research on prosody. Two of the speakers identified themselves as bilingual with Finnish as their stronger language. This was not considered a reason to exclude their data from the analysis, because familiarity with foreign languages is common in Finland, especially among students. Most of the subjects participating in Production Experiment I spoke several languages in addition to Finnish, as can be seen in Appendix A.[4]

Data had been collected from one further speaker (s01) who was not taken into account in the analysis. The participant generally seemed uncomfortable with the recording situation and used a speaking style differing strongly from the other participants, with flat contours and frequent prominent final rises (on the function of final rises in conversation, see Ogden and Routarinne, 2005, also see the discussion in sections 3.2.3 and 3.2.4 in the next chapter).

[4]One speaker, s11, declared to speak only one foreign language, English. However, he must at least have learnt Swedish in addition, as it is a compulsory subject in Finnish schools.

All participants were reimbursed for their time with a cinema ticket voucher.

2.2.3 Recording procedure

The 56 target sentences were interspersed with 110 filler sentences and presented to the participants on a computer screen one at a time using the programme Praat (Boersma and Weenink, 2010). A different pseudo-randomized order was employed for each speaker.[5] The subjects were instructed to first read the sentence silently. Then, pressing a button, they heard a pre-recorded question through loud-speakers and uttered the sentence as an answer to the question. The speakers were free to correct themselves and / or listen to the question again before proceeding to the next sentence by pressing another button.

The recordings were made in a recording studio at the department of Musicology of the University of Helsinki in May 2010, using a high-quality headset microphone, placed approximately 7 cm away from the speaker's mouth, with a sampling rate of 44,100 Hz.

2.2.4 Data editing, measurements and analysis

One sentence was accidentally skipped during the recording process and four other sentences were sorted out due to hesitations or slips of the tongue, so that altogether, $17 \times 56 - 5 = 947$ sentences and $947 \times 3 = 2841$ words were included in the analysis.

Acoustic measurements were done with the programme Praat (Boersma and Weenink, 2010). The recorded sentences were automatically segmented by Mikko Kurimo and Janne Pylkkönen with a forced alignment segmentation process developed at Aalto University (for a documentation of Aalto University's speech recognition programme using this process, see Hirsimäki et al., 2006, a demo can also be found online under `http://research.ics.aalto.fi/speech/wwwdemo.shtml`). The textgrids created based on these segmentations were automatically adjusted with a Praat script written by Antti Suni. Subsequently, I checked the segmentation manually, based on displays of waveform, spectrogram and auditory criteria. Then, for each word a Praat script identified the f_0 maximum (H) and the minima before (L1) and after the maximum (L2), if present. These detections were again manually verified to exclude effects of errors produced by Praat's pitch calculation algorithm; half of the sentences were checked by by me, the other half by a phonetically trained student with no knowledge of the objective of

[5]In all comparisons of statistical models, it was tested whether trial order, i.e. the position of an item in the recording session had an influence on the measured variable. Unless explicitly mentioned otherwise, no significant influence was found and the model was not improved by including trial order as a factor.

the study. At the same time, we marked stretches of speech that were articulated with non-modal voice quality. In doing so, we evaluated the auditory impression and inspected the oscillogram and spectrogram.

A second Praat script then measured the beginning, end and duration of all words, syllables and segments, as well as the f_0 and time of the marked f_0 turning points L1, H and L2. All measured f_0 values were converted to semitones (st) relative to 50 Hz for the analyses. The measurement script also automatically classified the tonal contour of the word as either 'LHL', 'HL', 'LH' or 'flat'. In addition to the automatic evaluation, I manually classified each word as carrying an early, mid or late peak or a contour that fit neither of these categories. This assessment was based on auditory evaluation as a first step and subsequent visual inspection as a second step.

Unless otherwise indicated, all statistical analyses were done by fitting linear mixed-effects models as implemented in R (R Development Core Team, 2010; Baayen, 2008). Different models were compared with the ANOVA function. Because these models are fairly complex, the maximum number of iterations was always set to 1000 (default is 300). Even with this limit, models with interactions in the random effects did usually not converge. Accordingly, when it is stated in model descriptions below that, for example, a model included random effects of quantity and position by subject this means that the model formula in R included the term `+(1+quantity+position|subject)`. In this example, the model estimations would assume that the factors quantity and position affected each subject to a different degree. For the binomial models of voice quality and pause occurrence (section 2.4.5 and 2.4.6), R's lmer function calculated p-values based on the z-scores. Otherwise, the calculation of p-values is not implemented for models with complex random factors. Therefore, the significance of effects was estimated from the t-values for all other models, treating an effect with $t > 2$ as significant. This should be fairly unproblematic for large data sets like the current one (Baayen et al., 2008, p. 398, footnote 1). Unless explicitly stated otherwise, the chapter will only discuss significant effects.

2.3 Hypotheses

The present chapter suggests that the data of the production experiment can be analysed in terms of prosodic phrasing. That is, it assumes that it is not necessary to account for the data by using contrasting accents. Instead it refers to p-phrases marked with H_pL_p tones and i-phrases aligned with focused constituents. In answer to the research questions **Qu1**, **Qu2** and **Qu3**, it thus advances three general hypotheses **H1**, **H2** and **H3** as given below. For the sake of clarity, the opposite assumptions regarding each question are given as possible alternative hypotheses **A1**, **A2** and **A3**.

Qu1 Can the data be accounted for by just one type of tonal specification on the phonological level or do they suggest a contrast between two or more accents?

- H1 There is tonal uniformity at the phonological level.
- A1 There are contrasting specifications (different types of accent).

Qu2 Is an analysis in terms of phrase tones or accents preferable for this data?

- H2 Phrase tones account better for the data.
- A2 Accents account better for the data.

Qu3 What is the shape and tonal specification of the contour(s)?

- H3 Contours are specified as falling (H_pL_p).
- A3 Contours are specified as rising-falling (LHL) or rising (L+H*).

The proposed phrasing analysis is exemplified in (11). It assumes that syntactic structure is taken as a basis for prosodic phrasing. Although the sentences are very short, an effect of position / grammatical function is thus anticipated: Since verb and object form a syntactic constituent, the verb phrase (VP), they should integrate into one p-phrase more often than subject and verb. In the absence of other influences, the sentence is also presumed to form a single i-phrase, creating the default prosodic structure in (11a).

(11) a. Broad focus:

$$\begin{array}{cccccccccc}
 & & H_p & & L_p & H_p & & L_p & L_i \\
 & & | & & | & | & & | & | \\
(& (& & \text{Manu} &)_p & (& \text{hali lelua} &)_p &)_i \\
 & & & \text{S} & & & \text{V}\quad\text{O} & &
\end{array}$$

'Manu hugged a toy.'

b. Narrow focus on the subject:

$$\begin{array}{ccccccccccc}
 & & H_p & & L_p & L_i & H_p & & L_p & L_i \\
 & & | & & | & | & | & & | & | \\
(& (& (& [\text{Manu}]_F &)_p &)_i & (& \text{hali lelua} &)_p &)_i
\end{array}$$

c. Narrow focus on the verb:

$$\begin{array}{cccccccccccccc}
 & & H_p & & L_p & H_p & & L_p & L_i & H_p & & L_p & L_i \\
 & & | & & | & | & & | & | & | & & | & | \\
(& (& (& \text{Manu} &)_p & (& [\text{hali}]_F &)_p &)_i & (& \text{lelua} &)_p &)_i
\end{array}$$

d. Narrow focus on the object: same phrasing as broad focus

A further hypothesis is that phrasing will be influenced by information structure, right-edge aligning focused constituents with i-phrases. Thus, an i-phrase boundary should be inserted after the narrow focus constituent in subject and verb focus, whereas no adjustment is necessary for narrow focus on the object, which is in final position. Although the proposal violates the strict layer hypothesis (Nespor and Vogel, 1986) by employing recursive phrasing on the i-phrase level (see the discussion in section 2.5.2), it does not assume level-skipping (Ito and Mester, 2009). That is, every phrase on level n dominates a phrase or phrases on level $n-1$. Thus, the insertion of an i-phrase boundary after a focused constituent implies the presence of a p-phrase boundary in the same location. As a result, (11) predicts that an SVO sentence with narrow subject focus differs from the broad focus version only on the i-phrase level, since the subject forms a p-phrase of its own in broad focus, too. Narrow verb focus, by contrast, should affect both levels, splitting up the p-phrase containing verb and object in broad focus. The third experimental variable, vowel quantity, is not expected to influence prosodic phrasing.

The following subsections will sketch the effects of the three experimental factors—information structure, position in the sentence and first syllable vowel quantity—that would be predicted for the six investigated parameters according to these overarching hypotheses and the results of previous research. Overall, strong interactions between the three factors can be expected. As was mentioned in section 2.2, the vowel quantity of each word's second syllable was determined by its grammatical function which in turn was correlated with its position. Also, the inherent correlation between information structure and word order and the special prosodic behaviour of the verbs have been mentioned. Although the experiment employed sentences with unmarked word order, an interaction between the two factors position and focus condition would thus not be surprising.

2.3.1 Tonal contour

The current data includes measurements for maximally three f_0 turning points per word (L1, H and L2), but not all three could be detected in all cases so that different tonal combinations arose. An automatic evaluation of the tonal contour of target words checked for the presence of three potential measurement points L1, H and L2 on each word, considering points differing from their neighbours by at least 0.8 st (see Appendix C on the choice and application of this criterion). As an illustration of the measurement points, consider Figure 2.1, marking them in red for the subject, in green for the verb and in blue for the object of an example sentence.[6] Since pitch is

[6]Notice that although analyses were done on semitone values, example pitch contours are displayed in Hz and are always smoothed using the smooth function in Praat (Boersma and Weenink, 2010) set at a bandwidth of 10 Hz.

rising throughout the sentence-medial verb, the f_0 maximum of this word is at its end and not followed by a local minimum word-internally. Thus, no L2 measurement point was identified for the verb. Its tonal contour was thus automatically classified as LH by the measurement script sorting word contours into the categories 'LHL', 'HL', 'LH' or 'flat' according to the presence or absence of the three possible measurement points.

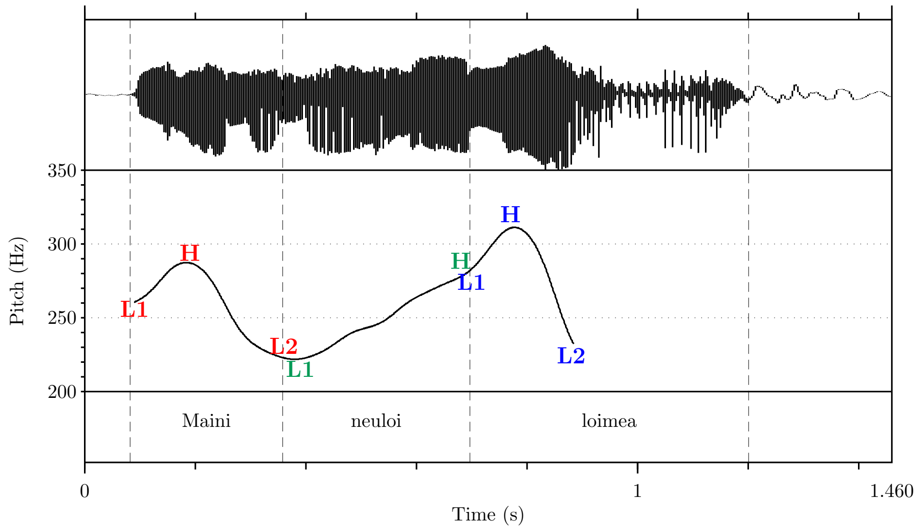

Figure 2.1: Illustration of measurement points. H = f_0 maximum, L1 = preceding minimum, L2 = following minimum of each word. Sentence *Maini neuloi loimea* 'Maini knitted a blanket', information focus on the object.

(12)
H_p L_p H_p L_p
Maini neuloi loimea
$(\text{Maini})_p$ $(\text{neuloi loimea})_p$
'Maini knitted a blanket.'

Note that this automatic classification into tonal contours is not a phonological analysis. Indeed, not all the measured points correspond to tonal targets under the present analysis in terms of phrase tones, given in (12). Instead, these points are defined objectively and independently of the analysis suggested here, namely as the f_0 maximum and minima of each word. These measurements provide a basis for the phonological account developed here and allow for a comparison with previous analyses assuming L+H* or rise-fall / LHL accents (Välimaa-Blum, 1993; Iivonen, 1998; Suomi et al., 2008). The additional subjective classification of pitch contours added an evaluation more focused on real f_0 turning points. Importantly, it recorded words with pitch contours falling from a pitch peak on the preceding word or rising to a peak on the following word, like the verb in Figure 2.1.

This chapter suggests that the tonal movements in the data from this experiment largely resulted from the phrase tones H_pL_p. It assumes that the measured L1 points do not correspond to phonological targets, whereas H and L2 are frequently, but not always, realisations of H_pL_p tones. In Figure 2.1, only H and L2 of the subject and object are analysed as corresponding to H_pL_p targets. By contrast, the points measured as L1 and H on the verb, as well as L1 on the object are modelled as an interpolation between the preceding L_p target associated with the p-phrase encompassing the subject and the following H_p tone realised on the object. The status of the sentence-initial L1 measurement point is left open at this point.

Regarding the three experimental factors, this account predicts effects of information structure and position / grammatical function, but not of vowel quantity. In general, the association of the H_pL_p with the p-phrase level predicts that they will not be realised on every word. As shown in (11) above, verbs and objects are expected to form a p-phrase together in broad focus and narrow object focus. Thus, a higher amount of incomplete LHL measurements are predicted for verbs, and possibly objects, in these information structural conditions, while subjects should have more complete LHL contours overall. In narrow verb focus, (11) hypothesises the insertion of an i-phrase boundary after the verb, making it a p-phrase of its own. A higher percentage of complete LHL measurements is thus predicted for narrow focus in position 2.

This account and the hypotheses it generates can be compared with other phonological analyses. The measurement of three f_0 turning points is in line with Suomi et al.'s (2008) assertion that Finnish pitch accents consist of three tonal targets, i.e. LHL (see Suomi et al., 2008, pp. 82, 122). Based on their description, one should thus expect that all three tones are realised on all words. The exception are unaccented words which "have no tonal properties of their own" (Suomi et al., 2008, p. 9). It can be predicted from their account that there will be no difference of tonal contour between words in broad focus, narrow information focus or narrow contrastive focus in the present material, as they describe all three accent types in their classification—thematic accents, rhematic accents and contrastive accents—as rise-falls (albeit with different levels of prominence, see Suomi et al., 2008, pp. 112–14).

The other bitonal description of Finnish accents by Välimaa-Blum (1988, 1993) models all basic tunes as consisting of either L+H* or (rarely) L*+H pitch accents, but the resulting contours are compatible with the descriptions by Suomi and his co-workers, the subsequent pitch fall being explained by the L target of the following word Välimaa-Blum (1993, p. 88). She specifically states that the same accent is used in narrow focus as in broad focus ('neutral contour') and also the unfocused words in the same sentence are labelled with

L+H* accents (see Välimaa-Blum, 1993, pp. 89, 93). On this basis, no effect of focus condition on the choice of tonal contour would therefore be predicted for the material discussed here.

An effect of grammatical function is expected from both accounts, as they describe finite verbs as generally unaccented. They do however imply slightly different predictions, as Suomi et al. (2008) state that these verbs are only accented in contrastive focus (see Suomi et al., 2008, p. 114), whereas Välimaa-Blum (1993) more generally speaks of (narrow) focus (see Välimaa-Blum, 1993, p. 84). No effect of quantity on the tonal contour has been suggested by Välimaa-Blum (1988, 1993) or Suomi et al. (2008).

The suggested analysis in terms of phrase tones and Välimaa-Blum's (1988; 1993) as well as Suomi et al.'s (2008) accounts share an important assumption: There is no tonal contrast, such as a difference between two or more accents, in Finnish utterances like those evaluated here. Instead, there is only one relevant tonal contour—phonologically specified as L+H* by Välimaa-Blum (1988, 1993), as LHL by Suomi et al. (2008) and as H_pL_p here—although the phrase tone account predicts that this contour can be spread across more than one prosodic word. However, a basis for an alternative hypothesis might be Vainio and Järvikivi's (2007) finding that the pitch fall is decisive in marking narrow focus in sentence-final position, whereas pre-finally, the pitch rise is more important. This could indicate a contrast between a falling and a rising accent. Complete LHL measurements in all positions would also be compatible with this description, but at least, one would not expect LH realisations in position three or HL realisations in position two.

2.3.2 Pitch scaling

Pitch scaling was evaluated for all measurable f_0 maxima (H) and preceding (L1) and following (L2) minima on each word. Of course, detecting all three points per word was not always possible, for example due to the presence of voiceless plosives or non-modal voice quality (see section 2.4.5). When only one point could be measured for a word, its f_0 was included among the measurements for H to give some indication about the word's pitch level.

An effect of information structure on pitch scaling is expected. Previous research has shown that pitch range is larger for narrow focus words in Finnish (e.g. Välimaa-Blum, 1993; Mixdorff et al., 2002; Vainio and Järvikivi, 2007). This is presumed to also affect the data of this chapter. More precisely, changes in pitch range are understood as stemming from differences in the scaling of tonal targets. Thus, the approach advanced in this thesis suggests that the scaling of the tonal targets H_p and L_p is affected. It is therefore hypothesised that H_p will be scaled higher and L_p lower in narrow focus, leading to higher H measurements and lower L2 measurements on words in narrow focus.

Regarding the effects of position, pitch usually displays a downward trend over the course of an utterances in Finnish (see e.g. Iivonen, 1998, p. 317), so lower measured f_0 values are expected later in the sentences of this chapter's data corpus. A direct effect of grammatical function is expected from previous research describing finite verbs as unaccented unless narrowly focused (Välimaa-Blum, 1993). Broad focus and unfocused verbs can be predicted to have smaller pitch ranges than their subject and object counterparts, while there should be no difference in the narrow focus condition.

As for quantity, no effect on pitch scaling is apprehended. In fact, Suomi et al. (2003) explicitly report the absence of differences between word structures in the size of both pitch rises and falls; this chapter is expected to reproduced this result.

2.3.3 Duration

A longer duration of words in narrow focus is hypothesised here on the basis of research on other languages, but also from what Suomi and his colleagues describe under the label 'accentual lengthening' (see e.g. Suomi et al., 2008, pp. 82–84.). However, they state that only 'contrastive', but not 'thematic' or 'rhematic' accents are realised with a longer duration (see Suomi et al., 2008, pp. 112–14), so that the prediction would not hold for narrow information focus. Thus, words in narrow information focus would be predicted to show no significant durational differences from those in broad focus or unfocused ones on the basis of this account.

Also, a difference between the positions should arise: Since (phrase or utterance) final lengthening has been attested in Finnish (see Suomi et al., 2008, pp. 108–109 for a summary), words in position 3 should have overall longer durations than those in positions 1 and 2.

Lastly, duration is known to be a marker of quantity in Finnish (see section 1.5). Obviously, the duration of quantity 2 vowels should be longer than that of quantity 1 vowels. Thus, the durations of words containing phonologically long vowels should be longer than the durations of words containing short vowels. More detailed predictions could be made and evaluated on the basis of previous studies, but since the the question of duration lies outside of the focal area of the present research topic, the matter will not be pursued further here. For this reason, section 2.4.3 will concentrate on the durations of complete words and not analyse syllable or segment durations.

2.3.4 Alignment

The present analysis predicts that the three experimental factors—information structure, position / grammatical function and vowel quantity—will not have a significant effect on tonal alignment. The central hypothesis regarding tonal alignment is that the data in this chapter will not exhibit a distinction between two or more pitch accents. In contrast, this thesis suggests an account in terms of phrase tones, the relevant tones here being H_p and L_p. As depicted in (11) above, this analysis assumes that these tones are associated with the left and right edge of the p-phrase, respectively. Thus, H_p is predicted to appear early in the p-phrase and L_p should appear towards its end. The precise alignment is left open at this point. Crucially, while some variation in the realisation of these tones is expected, no systematic differences between the experimental conditions should occur based on this account.

As related in Chapter 1.4.3, there are two bitonal accounts characterising Finnish intonation in terms of accents. Both generally agree with the hypothesis of a uniform phonological specification, but differ slightly from each other. According to Välimaa-Blum (1988, 1993), the near-universally used Finnish accent is L+H*, with the H tone realised "on or in the immediate vicinity of the primary stressed syllable" (Välimaa-Blum, 1988, p. 104), usually at the end of the stressed syllable (Välimaa-Blum, 1993, p. 88). The L tone was localised at the beginning of the stressed syllable in her data (see ibid.). Additionally, Välimaa-Blum (1993) mentions an L*+H accent, although the licensing context or function of this accent is not clear. In this case, a later alignment of the H tone would be expected, with the L tone keeping the pitch of the stressed syllable low and the H following it "after some, probably specific, time delay" (Välimaa-Blum, 1993, p. 88).

Suomi and his colleagues specify that the first L tone of their LHL accents is aligned to the onset consonant, the H to the end of the first mora and the second L tone a bit more imprecisely around the end of the second mora or the middle of the third syllable (Suomi, 2007a; Suomi et al., 2008). The reference to moraic structure means that when the first syllable contains two moras, H tones should always be realised well within the syllable nucleus, whereas for monomoraic first syllables realisations are expected at the end of the first syllable nucleus or during second syllable's onset consonant.

A specific implication of the phonological uniformity presumed here regards quantity. It is predicted that the induced variation in (first syllable vowel) quantity does not affect the timing of the peak (H). Instead, pitch peaks are expected to occur at the same distance from the beginning of the word for quantity 1 as for quantity 2 (as in studies reported by Wiik, 1988; Suomi, 2007a). Possibly, an alternative hypothesis can be derived from Vainio et al.'s (2010) study. Employing Xu's (1999) target approximation model (Xu, 1999, 2005), they describe finding a high static tonal target on initial quantity 1 syllables and a dynamic falling target for initial quantity 2 syllables. The target of the second syllable is a static L irrespective of its quantity in this account. In the AM framework, both cases should probably be described as realisations of HL tones, but it could be speculated that the peak H occurs towards the end of the first syllable for quantity 1, but further towards its beginning for quantity 2.

2.3.5 Voice quality

A phenomenon that is easy to discern also in everyday conversations is the usage of creaky or breathy voice as a finality marker (for a short summary of the literature see Suomi et al., 2008, p. 137–139). This leads to the clear hypothesis that non-modal voice quality should occur more frequently in sentence-final words than in position one and two.

Thus, non-modal voice is presumed to mark the ends of prosodic phrases, but it is so far not clear where in the prosodic hierarchy this phenomenon is located (i.e. whether it marks the end of p-phrases, i-phrases or even smaller or larger units). Thus, while information structure is presumed to affect prosodic phrasing and prosodic phrasing is presumed to affect the occurrence of non-modal voice quality, concrete predictions regarding the effect of focus condition cannot be derived at this point.

There are no concrete predictions regarding effects of quantity on voice quality at this point.

2.3.6 Occurrence of pauses

The sentences elicited for this chapter were short and syntactically simple. It is therefore plausible that speakers rarely punctuate these sentences with pauses. Thus, an overall low number of pauses is predicted here. The distribution of any occurring pauses is tentatively hypothesised to differ between positions / grammatical functions and focus conditions, since syntax and information structure are presumed to influence prosodic phrasing, as exemplified in (11). Pauses are expected to occur more frequently at the boundaries of larger phrases. According to (11), an i-phrase boundary appears after subjects and verbs in narrow focus, so that more pauses are predicted in these locations in these information structural conditions. By contrast, an effect of vowel quantity is not expected.

2.4 Results

2.4.1 Tonal contour

This section gives the results of two evaluations of tonal contour for each word, hypothesized to be influenced by information structure and differing between positions in the sentence. First, it reports the presence of the three automatically identified and manually checked f_0 turning points for all the words—the maximum (H) and the minima left (L1) and right (L2) of it. Second, it presents the results of a subjective assessment of word contours. The data show that words frequently, but not always were realised with rise-fall contours of their own, with more complete rise-falls on words in narrow focus and less on sentence-medial verbs.

For 324 words in the current data set, no f_0 values or only an extremely short stretch of the f_0 contour could be measured due to non-modal voice quality, so that these words could not be automatically classified in terms of their tonal contour. Among the rest, a maximum and two minima left and right of it could be identified for most words, so that by far the most frequent category was LHL. Also the combinations HL (f_0 measurement points H and L2) and LH (L1 and H) occurred. Lastly, there was a small number of completely flat contours (that is, words for which none of the measured values differed from the neighbouring ones within the same word by more than 0.8 st, see Appendix C).

Table 2.2 shows the distribution of automatic classifications by information structural condition and position in the sentence. Within each condition, the classifications as LHL constitute the largest group, but differences can still be found. Compared to broad focus, the percentage of LHL classifications was higher in narrow focus and lower on unfocused words. While all three measurement points were detected on between 83% and 99% of narrowly focused words, this was only the case for about a third of the sentence-medial unfocused words. Thereby, numbers were very similar for contrastive and information focus. Also for the unfocused given words there was no difference between the two focus types, i.e. it did not matter whether the narrow focus constituent occurring in the same sentence was in narrow information or contrastive focus. Conversely, HL and LH realisations each appeared on about a fifth of unfocused and broad focus words overall, but only for about 5% of narrow focus words. A seizable number of flat tonal contours only appeared on unfocused words.

A difference between sentence-initial, -medial and -final position also emerged from the data, although LHL was again clearly the largest category in all conditions. As illustrated by Table 2.2, the smallest proportion of complete measurable LHL contours appeared in position 2. However, even in unfocused condition, LHL was still the most frequent combination of tones, found on about a third of the sentence-medial verbs. Additionally, Table 2.2 exhibits at least two interesting connections between the factors focus condition and position. First, measurable contours on sentence-initial words were less strongly influenced by focus condition than those on words in the other two positions. More LHL categorisations in narrow focus still appeared in position 1, but there was no difference in the frequency of LHL between broad focus and unfocused subjects. Second, the combinations HL and LH were differently distributed by position in unfocused condition: While HL appeared more frequently on unfocused sentence-initial subjects, LH measurement points were detected more frequently on unfocused objects. For position 2, no clear preference emerges from Table 2.2, but a closer look revealed an influence of focus location. In sentences with narrow (information or corrective) focus on the subject, unfocused verbs were marked as HL in 38% of the cases and as LH in 16% (N = 70 and N = 29, respectively). In

Table 2.2: Number of occurrences and percentage (in brackets, rounded) for combinations of f_0 measurement points by focus conditions and position. CF = contrastive focus, IF = information focus. For unfocused given words, the table additionally specifies in brackets which type of focus was present for the focused constituent of the same sentence.

		flat	HL	LH	LHL
Broad focus	Position 1	1	23	21	89
		(1%)	(17%)	(16%)	(66%)
	Position 2	13	22	43	55
		(10%)	(17%)	(32%)	(41%)
	Position 3	3	26	6	94
		(2%)	(20%)	(5%)	(73%)
Broad focus	*Sum*	17	71	70	238
	%	4	18	18	60
Narrow CF	Position 1	0	9	3	122
		(0%)	(7%)	(2%)	(91%)
	Position 2	0	9	13	114
		(0%)	(7%)	(10%)	(84%)
	Position 3	0	1	1	133
		(0%)	(1%)	(1%)	(99%)
Narrow IF	Position 1	0	15	8	111
		(0%)	(11%)	(6%)	(83%)
	Position 2	1	10	11	113
		(1%)	(7%)	(8%)	(84%)
	Position 3	0	3	1	132
		(0%)	(2%)	(1%)	(97%)
Narrow focus	*Sum*	1	47	37	725
	%	0	6	5	90
Unfocused (CF)	Position 1	13	55	23	177
		(5%)	(21%)	(9%)	(66%)
	Position 2	25	70	54	76
		(11%)	(31%)	(24%)	(34%)
	Position 3	22	21	50	50
		(15%)	(15%)	(35%)	(35%)
Unfocused (IF)	Position 1	15	55	18	180
		(6%)	(21%)	(7%)	(67%)
	Position 2	25	60	70	73
		(11%)	(26%)	(31%)	(32%)
	Position 3	32	26	59	62
		(18%)	(15%)	(33%)	(35%)
Unfocused	*Sum*	*132*	*287*	*274*	*618*
	%	10	22	21	47

contrast, 22% of unfocused verbs in object focus sentences were categorised as HL and 35% as LH (N = 60 and N = 95, respectively). In narrow focus, HL and LH were about equally infrequent in all positions.

Finally, an evaluation in terms of first syllable vowel quantity found no noteworthy difference between the two quantity conditions. The LHL contour occurred on 759 and 822 words or 62% and 64% for quantity 1 and 2, respectively, while HL appeared for 222 and 183 words (18% and 14%), LH 171 and 210 words (14% and 16%) and 71 and 79 words were identified as having flat f_0 (6% for both quantities).

The subjective classification of tonal contours divided words into those with early, mid and late peaks, based on auditory and visual impression. Of these, the largest group were the early peaked contours, which amounted to 45% of all evaluated words (N = 1280). Some words did not fit these three categories, but were classified as not carrying a contour of their own, whereby three cases were distinguished. First, altogether 232 words (8%) were perceived as having flat pitch, i.e. no perceivable pitch movements. Second, the f_0 contour of some words was defined by a pitch fall from the preceding word's peak, as illustrated by the sentence-medial verb in Figure 2.2. Third, the reverse case also occurred, so that a word already carried a part of the the rise to the following word (see Figure 2.3, already discussed in section 2.3.1). Together, these cases amounted to 270 words or just under 10% of the data. Finally, a category 'other' comprised 471 words (17%) with prevailing non-modal voice or otherwise not fitting any of the other labels.

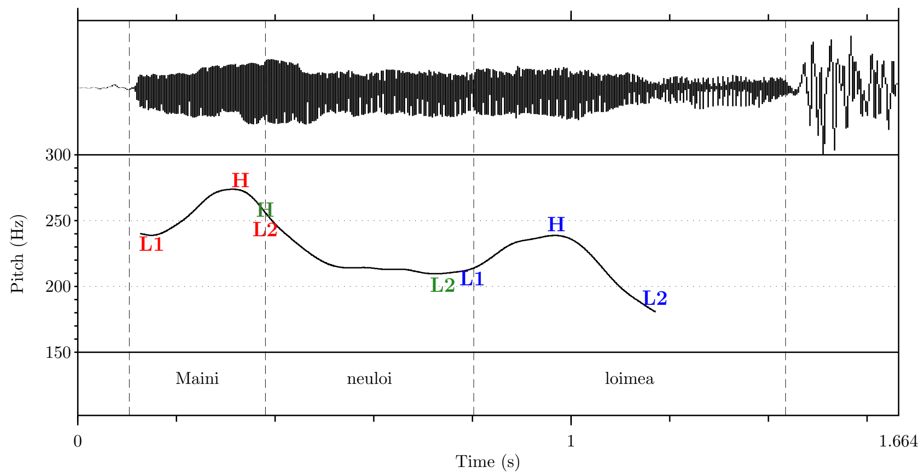

Figure 2.2: Pitch movement on the verb is a fall from the subject peak (followed by a plateau). Sentence *Maini neuloi loimea* 'Maini knitted a blanket', information focus on the object.

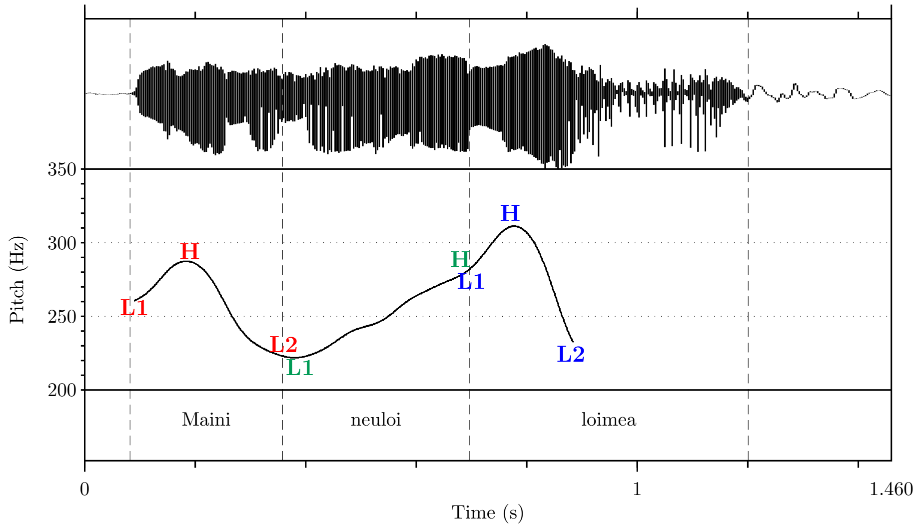

Figure 2.3: Pitch movement on the verb is a rise to the object peak. Sentence *Maini neuloi loimea* 'Maini knitted a blanket', information focus on the object (= Figure 2.1).

Although these subjective categorisations were based on my non-native judgement, there was a good correspondence between focus condition and impressionistic classification in line with the automatic evaluation discussed above. As shown in Table 2.3, the differences between information and contrastive focus were again small. Most words in narrow focus were judged to carry an early peak, while this was the case for a little less than half of the broad focus words. Narrow focus words additionally were perceived to have mid peaks more often than broad focus ones. The category 'late' was used rarely altogether, but even less in narrow focus than in the other focus conditions. In contrast, unfocused words carried the smallest percentage of early or mid peaks, while being categorised as flat or 'other' most often. They also sometimes carried a part of the preceding or following word's contour as their only significant pitch movement, although slightly less often than words in broad focus. The 'fall from' and 'rise to' contours practically did not occur on narrow focus words at all, which were always perceived to have—early, mid or late peaked—tonal contours of their own.

Table 2.4 gives the impressionistic categorisations for the different sentence positions or subjects, verbs and objects, respectively. In all positions, peaks were most frequently perceived as early. However, differences also occurred. For instance, the category 'other' was used for about a third of the words in position 3, which can be connected to the use of sentence-final non-modal voice hampering the judgement of accent type (see section 2.4.5). In contrast, the category 'other' was almost never chosen for sentence-initial

Table 2.3: Number of occurrences and percentage (in brackets, rounded) for perceived contour types by focus conditions. CF = contrastive focus, IF = information focus. For unfocused given words, the type of narrow focus present elsewhere in the sentence is also specified.

	early	mid	late	flat	fall from preceding	rise to following	other
Broad focus	195 (48%)	55 (14%)	26 (6%)	29 (7%)	30 (7%)	49 (12%)	21 (5%)
Narrow CF	282 (70%)	113 (28%)	10 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Narrow IF	306 (75%)	89 (22%)	7 (2%)	0 (0%)	2 (0%)	1 (0%)	2 (0%)
Unfocused (CF)	227 (28%)	83 (10%)	60 (7%)	95 (12%)	38 (5%)	51 (6%)	256 (32%)
Unfocused (IF)	270 (33%)	87 (11%)	58 (7%)	108 (13%)	33 (4%)	66 (8%)	192 (24%)

Table 2.4: Number of occurrences and percentage (in brackets, rounded) for perceived contour types by position of the word in the sentence.

	early	mid	late	flat	fall from preceding	rise to following	other
Position 1	465 (49%)	317 (33%)	109 (12%)	14 (1%)	0 (0%)	35 (4%)	7 (1%)
Position 2	319 (34%)	85 (9%)	43 (5%)	134 (14%)	81 (9%)	132 (14%)	153 (16%)
Position 3	496 (52%)	25 (3%)	9 (1%)	84 (9%)	22 (2%)	0 (0%)	311 (33%)

subjects, where the 'mid' category was in turn more frequent than later in the sentence. Also, both categorisations as 'fall from' and as 'rise to' contours emerged most often in position 2. In fact, verbs and objects frequently shared one rise-fall contour. As examples, consider Figures 2.4 and 2.5. Figure 2.4 looks similar to the narrow object focus example in Figure 2.3, but shows a downstep of the object's f_0 maximum relative to the maximum of the subject. In Figure 2.5, the two constituents of the VP likewise share a rise-fall contour, but the peak is located at on the verb. Both examples were realised in broad focus contexts. In informal experiments with manipulated stimuli, three native speakers consistently judged contours of both types to be good answers to questions inducing broad focus (all-new) or VP-focus, but not as inducing a perception of narrow focus on either of the constituents.

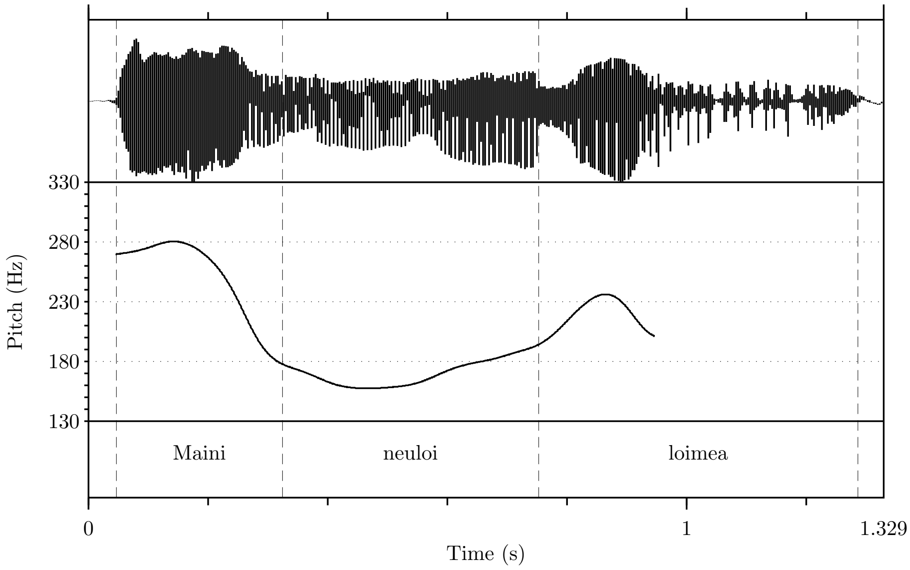

Figure 2.4: Verb and object sharing a rise-fall pitch contour with the peak realised on the object. Sentence *Maini neuloi loimea* 'Maini knitted a blanket', broad focus.

Finally, a noteworthy difference in the author's classifications was also found regarding the factor quantity: Peaks were perceived as mid and late more often for words with a short first syllable vowel than for those with a long one (see Table 2.5). Notice, however, that like the effects of the other factors, also these differences in the subjective judgements did not indicate a shift in preference for a certain category in a certain condition, but rather a more or less strong prevalence of the 'early peak' classification.

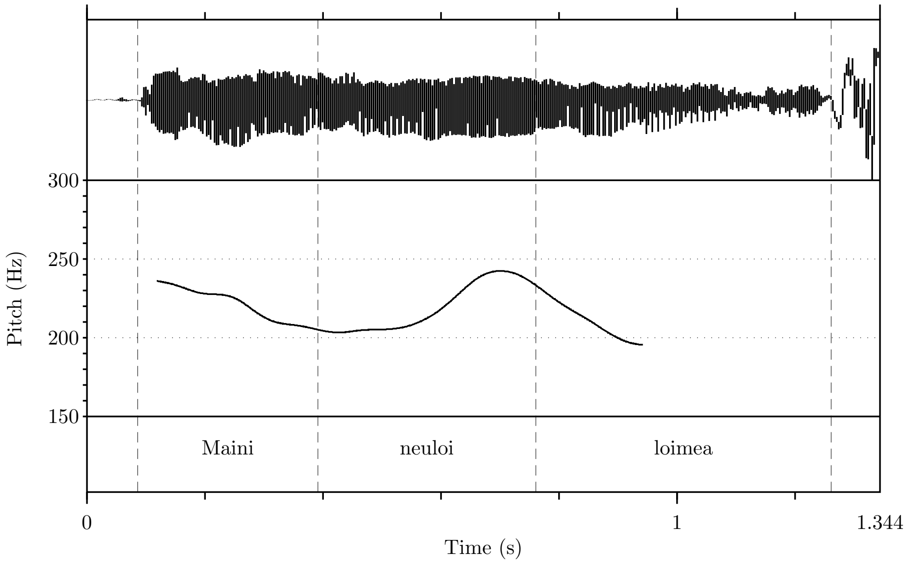

Figure 2.5: Verb and object sharing a rise-fall pitch contour with the peak realised on the verb. Sentence *Maini neuloi loimea* 'Maini knitted a blanket', broad focus.

Table 2.5: Number of occurrences and percentage (in brackets, rounded) for perceived contour types by first syllable vowel quantity.

	early	mid	late	flat	fall from preceding	rise to following	other
Quantity 1	565 (40%)	278 (20%)	114 (8%)	104 (7%)	52 (4%)	91 (6%)	221 (16%)
Quantity 2	715 (50%)	149 (11%)	47 (3%)	128 (9%)	51 (4%)	76 (5%)	250 (18%)

2.4.2 Pitch scaling

This section reports the scaling of the three automatically identified and manually checked f_0 turning points for all the words—the maximum (H) and the minima left (L1) and right (L2) of it. It will look at the effects of information structure, position and quantity in turn before investigating the correlations between the three measurement points. Altogether, 2081 values were evaluated for L1, 2533 for H and 2245 for L2 here. Section 2.3.2 predicted an effect of focus condition on the phonological targets H_p and L_p, generally corresponding to H and L2. Also, it anticipated a smaller pitch range for broad focus and unfocused verbs than for other constituents in the same focus conditions. No effect of quantity was anticipated.

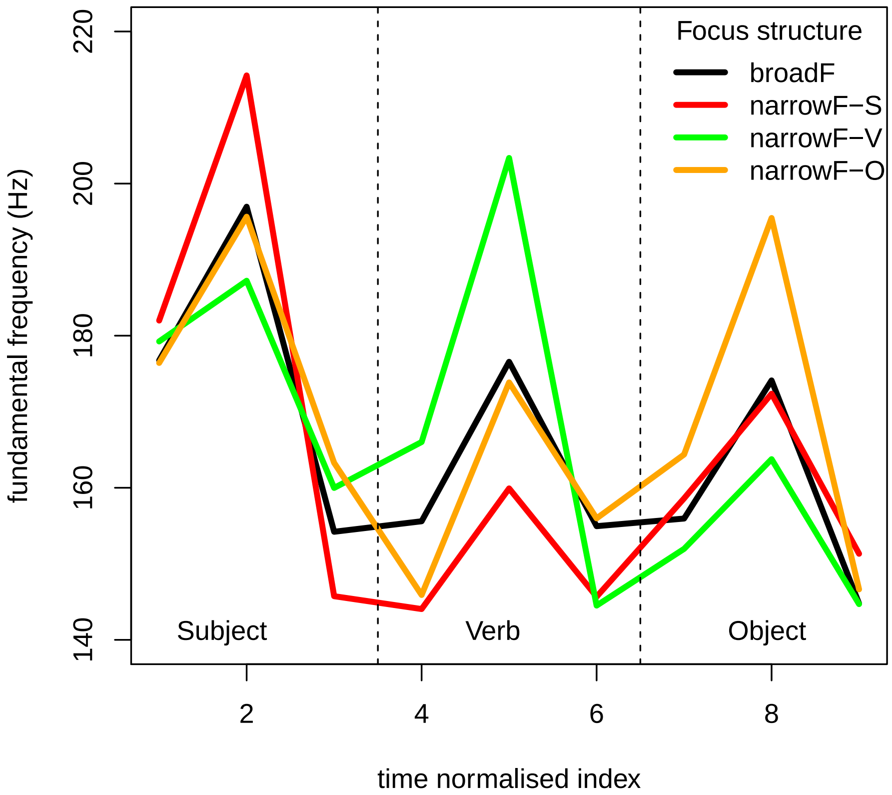

Figure 2.6: Time-normalised average sentence contours from three pitch values per word for broad vs. narrow focus with different focus locations.

Figure 2.6 illustrates the mean f_0 for the three turning points for each constituent of the sentence—in other words average sentence contours—in different focus structures. Thereby, the graph separates broad focus and narrow focus in each of the three locations, but for the sake of clarity does not distinguish between information focus and contrastive focus.[7] It shows that narrow focus affected the pitch scaling of minima and maxima in all three positions. The most apparent effect is that on the peaks (H), which were realised much higher in narrow focus than in broad focus, but lowered on unfocused given words. These effects also appeared in the linear mixed-effects model in Table 2.6, containing positive estimates for the effects of narrow information focus and narrow contrastive focus. This indicates that the higher peaks in narrow focus were significant compared to the broad focus intercept—an effect that was even stronger in the second and third position, as the interactions indicate (one of the these interactions, between information focus and the sentence-final position, had a t-value of 1.84 which is interpreted as only marginally significant here). The lowering in unfocused condition did not constitute a significant main effect, but interactions showed that pitch maxima of unfocused words were significantly lower in sentence-final position. Marginal interactions for position 2 pointed into the same direction.

Table 2.7 presents the best model of the first pitch minimum (L1). Like the pitch maxima, L1 values were realised significantly higher in narrow focus, as suggested by positive main effects of both types of narrow focus, visible in Figure 2.6. L1 was also raised for unfocused given words, but this tendency reached significance as a main effect only for information focus. There were, however, significant interactions of the unfocused condition with both position and quantity: L1 values were lower for unfocused sentence-medial and sentence-final words. L1 also had a lower f_0 in unfocused quantity 2 words.

While the peaks (H) and preceding minima (L1) were realized significantly higher in narrow focus than in broad focus condition, the following valleys (L2) were scaled lower, as confirmed by the model in Table 2.8. Thus, narrow focus words exhibited an increased pitch range, while unfocused words had smaller pitch ranges due to lower H and higher L2 values (see Table 2.8). Figure 2.7 illustrates the these two effects and their interpretation under the current analysis: In this narrow verb focus example, each word constitutes a p-phrase of its own and is thus marked with $\mathrm{H_pL_p}$ tones. These tones are further apart for the narrow focus verb and closer together for the pre- and post-focal constituents. A comparison with the contours in Figure 2.6 indicates that pitch range adjustments were not always this ex-

[7]The difference between the two narrow focus types is however not very large, as indicated by the linear mixed-effects models which usually show the same effects for information focus and contrastive focus, albeit not always the same effect size (see Tables 2.6–2.8).

Table 2.6: Coefficients of the best linear mixed-effects model of f_0 of H (in st) with random effects of item and of focus condition and position by subject. CF indicates contrastive focus, IF information focus on the on the narrow focus constituent.

	Estimate	Std. Error	t value
(Intercept)	27.2690	0.6272	43.4763
Narrow focus (CF)	1.7009	0.3540	4.8047
Narrow focus (IF)	1.4875	0.3156	4.7135
Unfocused (CF)	−0.3041	0.3266	−0.9312
Unfocused (IF)	0.1087	0.2991	0.3636
Position 2	−1.7654	0.3708	−4.7611
Position 3	−1.9741	0.4308	−4.5822
Quantity 2	0.7374	0.2052	3.5940
Male gender	−12.1901	0.6602	−18.4633
Narrow focus (CF) : position 2	1.1954	0.2836	4.2155
Narrow focus (IF) : position 2	0.8427	0.2836	2.9717
Unfocused (CF) : position 2	−0.4476	0.2498	−1.7916
Unfocused (IF) : position 2	−0.4559	0.2493	−1.8286
Narrow focus (CF) : position 3	0.8389	0.2853	2.9403
Narrow focus (IF) : position 3	0.5231	0.2848	1.8372
Unfocused (CF) : position 3	−0.9943	0.2677	−3.7141
Unfocused (IF) : position 3	−1.2277	0.2593	−4.7356
Narrow focus (CF) : quantity 2	−0.6207	0.2324	−2.6705
Narrow focus (IF) : quantity 2	−0.5014	0.2323	−2.1589
Unfocused (CF): quantity 2	−0.9099	0.2112	−4.3092
Unfocused (IF) : quantity 2	−0.9529	0.2082	−4.5776

Table 2.7: Coefficients of the best linear mixed-effects model of f_0 of L1 (in st) with random effects of focus condition by item and of focus condition, (first syllable) vowel quantity and position by subject. CF indicates contrastive focus, IF information focus on the on the narrow focus constituent.

	Estimate	Std. Error	t value
(Intercept)	24.8149	0.5538	44.8059
Narrow focus (CF)	1.1488	0.4208	2.7298
Narrow focus (IF)	1.4300	0.4543	3.1480
Unfocused (CF)	0.4580	0.3251	1.4086
Unfocused (IF)	0.9025	0.3091	2.9201
Position 2	−1.4308	0.4377	−3.2691
Position 3	−1.3218	0.4371	−3.0240
Quantity 2	0.4556	0.2801	1.6265
Male gender	−11.9003	0.5762	−20.6537
Narrow focus (CF) : position 2	0.1570	0.4920	0.3190
Narrow focus (IF) : position 2	−0.2595	0.5146	−0.5042
Unfocused (CF) : position 2	−0.8885	0.3810	−2.3320
Unfocused (IF): position 2	−0.7921	0.3599	−2.2008
Narrow focus (CF) : position 3	−0.3461	0.4921	−0.7032
Narrow focus (IF) : position 3	−0.1647	0.5145	−0.3201
Unfocused (CF) : position 3	−1.6870	0.3902	−4.3233
Unfocused (IF) : position 3	−1.8974	0.3663	−5.1799
Narrow focus (CF) : quantity 2	−0.5697	0.4021	−1.4169
Narrow focus (IF) : quantity 2	−0.4008	0.4200	−0.9542
Unfocused (CF) : quantity 2	−0.7000	0.3157	−2.2171
Unfocused (IF) : quantity 2	−0.7368	0.2978	−2.4744

Table 2.8: Coefficients of the best linear mixed-effects model of f_0 of L2 (in st) with random effects of item and of focus condition and position by subject. CF indicates contrastive focus, IF information focus on the on the narrow focus constituent.

	Estimate	Std. Error	t value
(Intercept)	23.1146	0.5865	39.4079
Narrow focus (CF)	−0.8084	0.3061	−2.6409
Narrow focus (IF)	−0.6952	0.2830	−2.4561
Unfocused (CF)	0.6860	0.2545	2.6952
Unfocused (IF)	0.9337	0.2456	3.8013
Position 2	0.2501	0.3790	0.6597
Position 3	−1.5117	0.4548	−3.3242
Quantity 2	−0.0270	0.2554	−0.1057
Male gender	−10.8450	0.6816	−15.9117
Narrow focus (CF) : position 2	−0.0017	0.3090	−0.0054
Narrow focus (IF) : position 2	−0.0997	0.3108	−0.3208
Unfocused (CF) : position 2	−1.3149	0.2798	−4.6997
Unfocused (IF) : position 2	−1.0065	0.2804	−3.5890
Narrow focus (CF) : position 3	1.6217	0.3016	5.3771
Narrow focus (IF) : position 3	1.4887	0.3019	4.9312
Unfocused (CF) : position 3	−1.1022	0.3047	−3.6168
Unfocused (IF) : position 3	−0.7073	0.2848	−2.4831
Narrow focus (CF): quantity 2	−1.1093	0.2494	−4.4483
Narrow focus (IF) : quantity 2	−0.7494	0.2499	−2.9989
Unfocused (CF) : quantity 2	−0.0979	0.2362	−0.4144
Unfocused (IF) : quantity 2	−0.4658	0.2317	−2.0100
Position 2 : quantity 2	−0.4660	0.2697	−1.7278
Position 3 : quantity 2	0.5231	0.2766	1.8913

tensive. However, they were pervasive enough to appear clearly in the mean values and to reach significance in statistical modelling.

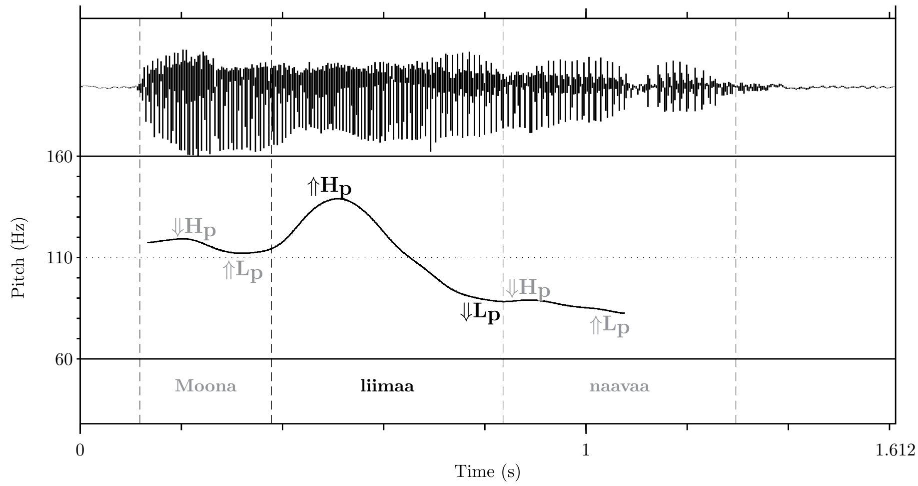

Figure 2.7: Pitch range expansion on a word in narrow focus (black) and pitch range compression on unfocused words (grey). Sentence *Moona liimaa naavaa* 'Moona glues lichen', narrow contrastive focus on the verb.

In addition to the influence of information structure, pitch scaling was affected by the factor position, as can also be seen in Figure 2.6. A general downtrend of pitch peak values over the course of the sentence can be observed, and a declination of the top-line is visible even from the comparison of narrow focus peaks. In keeping with this, the linear mixed-effects model of measured H values (see Table 2.6) showed a main effects of position, indicating that pitch maxima were overall lower in position 2 and position 3 than sentence-initially. The size of this effect was only slightly larger for sentence-final objects than for sentence-medial verbs. Thus, the drop in peak f_0 relative to the previous constituent was larger for verbs than for objects—as is especially visible from the broad focus sentence contour in Figure 2.6. Also for the scaling of L1, a main effect of position in Table 2.7 indicates that the measured f_0 was overall significantly lower later in the sentence, pointing to a general pitch downtrend. In contrast, the f_0 values measured as L2 were overall lower in position 3 than for the position 1 intercept, but the effect was not significant for position 2. However, there were several interactions affecting the scaling of L2 for sentence-final words. As Figure 2.6 illustrates, they virtually converged on a low point in all focus conditions. Thus, both the lowering of L2 in narrow focus and the raising of L2 in unfocused condition had a smaller effect in position 3, as reflected by two significant interactions between the factors information structure and position in the model in Table 2.8. The model also included interactions indicating that L2 values of unfocused words were significantly lower in position 2. Figure 2.6 suggests that this is mainly due to the narrow subject focus condition (red line), where pitch on the—post-focally given—verb was overall very low.

Vowel quantity also played a certain role in pitch scaling. For the scaling of peaks, a significant main effect of quantity appeared in the model in Table 2.6. It estimated pitch maxima to be overall slightly higher in quantity 2 than in quantity 1. Note however, that all the interactions of quantity with other factors show a negative estimate in the model, counteracting the main effect. For the scaling of L1, there was no main effect of quantity, but an interaction with the unfocused condition (see Table 2.7). An otherwise identical model of L1 values excluding the factor quantity (from fixed effects, random effects or both) did not converge, so that it is not possible to say whether this factor made a significant contribution to the model. For the scaling of L2, the interactions between the narrow focus conditions and quantity 2 showed a boosting of the lowering effect. That is, quantity 2 words had lower L2 values in narrow focus than quantity 1 words, as shown by two interactions in Table 2.8.

Finally, a significant effect of gender emerged for all three models in Tables 2.6–2.8: On average, men have lower voices than women due to longer vocal cords. Therefore, lower f_0 values were expected and found for male speakers at all three measurement points (L1, H and L2) here.

In summary, the results for the scaling of L1 and H were overall relatively similar, displaying raising in narrow focus and lowering in unfocused condition, while L2 showed the opposite behaviour with respect to focus condition. It can therefore be asked whether L1 and H within each word were more similar to one another than H and L2 within the same word. Indeed the comparison in Figure 2.8 shows that the correlation between L1 and H was stronger than that between L2 and H, with the lowess (locally weighted scatterplot smoothing) curve fit to the data approaching a linear line more closely in the top panel.

Transgressing the unit of a single word, Figure 2.9 shows the correlation between the two pairs of neighbouring low tones measured in a sentence—i.e. between L2 of the subject and L1 of the verb (top panel) and between L2 of the verb and L1 of the object (bottom panel)—in different focus conditions. Again, the measurements from male and female speakers formed two distinguishable groups, but even though they were depicted in the same plot, the lowess curves are almost perfectly linear with a slope around 45°. This suggests that the L2 values of subjects were very similar to the L1 values of verbs and that L2 of verbs and L1 of objects were also very close to each other.

Indeed, it is difficult to discern the realisation of two separate targets from many of the pitch contours. For example, L2 of the verb and L1 of the object are close together in Figure 2.10, where the rise to the object

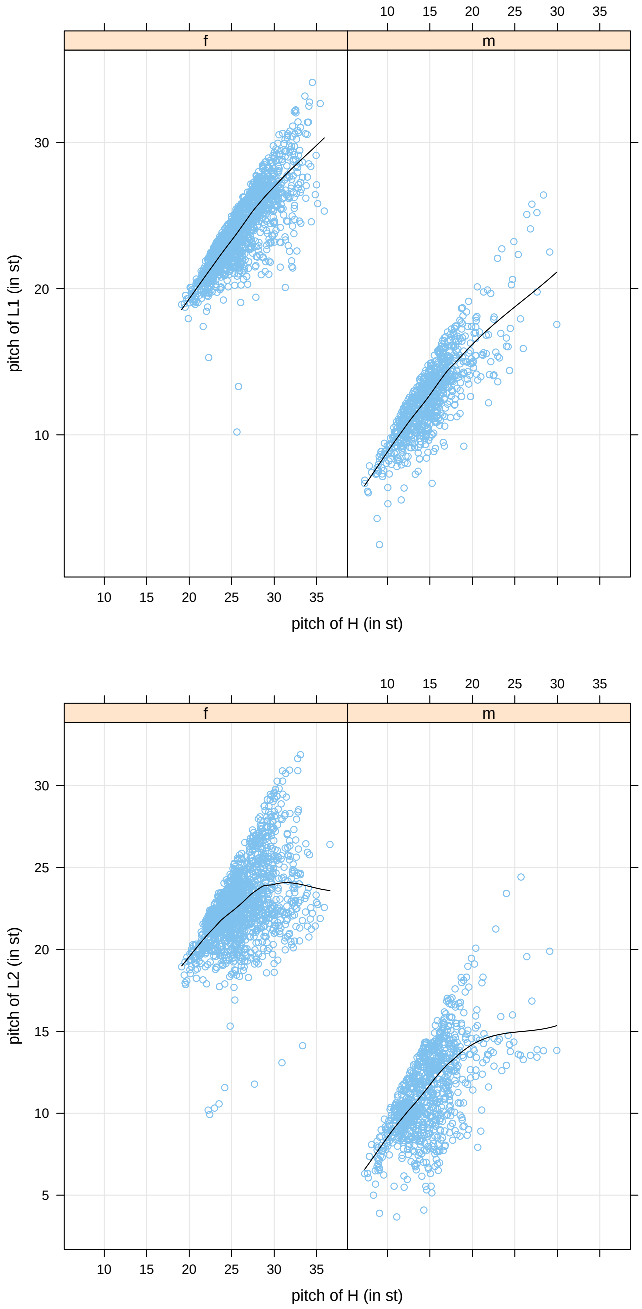

Figure 2.8: Correlation between pitch of measured low and high f_0 values within a word. Top panel: Pitch (in st) of L1 as a function of pitch of H for female (left) and male (right) speakers. Bottom panel: Pitch (in st) of L2 as a function of pitch of H for female (left) and male (right) speakers.

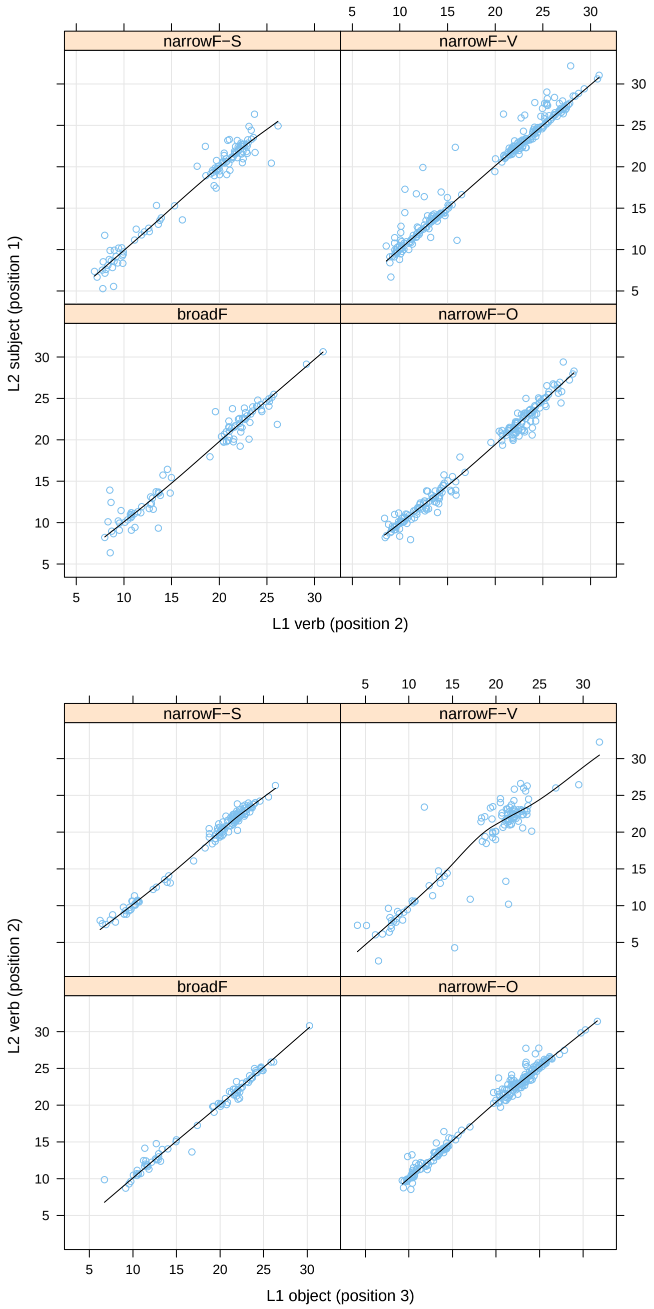

Figure 2.9: Correlation between neighbouring low values within a sentence for different information structural conditions. Top panel: L1 of verb as a function of preceding low (L2 of subject). Bottom panel: L1 of object as a function of preceding low (L2 of verb).

peak already starts on the verb (likewise, the rise to the verb peak starts on the subject). A quantitative impression can be gleaned by subtracting the values of two points identified as neighbouring L2 and L1. The mean absolute difference between L2 of subjects and L1 verbs was 0.72 st (6 Hz) and the median absolute difference 0.34 st (3 Hz). For L2 of verbs and L1 of objects, the respective mean value was 0.71 st (6 Hz) and the median 0.38 st (3 Hz). Temporally, the points were sometimes very close together, with minimum distances of just above one millisecond (ms), but measured as separate points due to the criteria of marking tonal points. There was, however, a large variation, so that the mean distance between the subject's L2 and the verb's L1 was 75 ms and the median 39 ms, while the mean distance between the verb's L2 and the object's L1 was 119 ms and the median 85 ms.

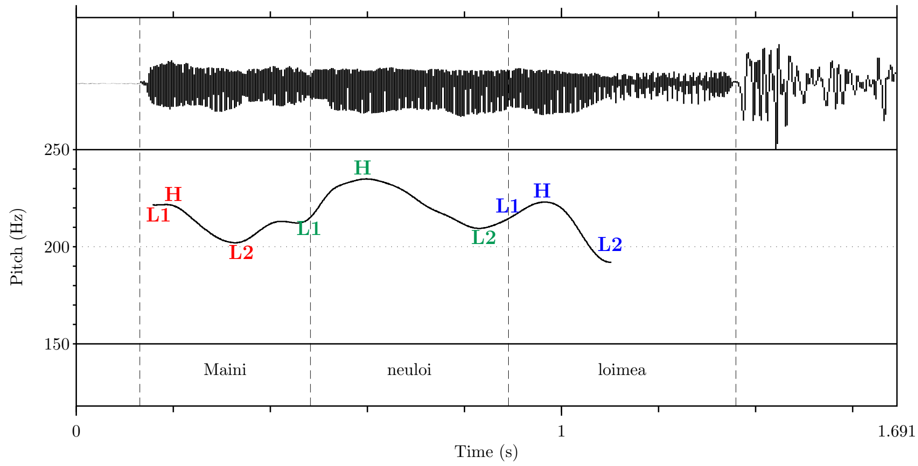

Figure 2.10: Measurement points L2 for the verb and L1 for the object do not differ clearly. Sentence *Maini neuloi loimea* 'Maini knitted a blanket', broad focus.

2.4.3 Duration

This section evaluates word duration, testing the hypotheses that durations were longer in narrow focus, position 3 and for words containing quantity 2 vowels. The mean durations of words and segments in the current data are depicted in Figure 2.11, where dark grey bars depict the average durations of consonants and lighter grey bars those of vowels. The panels distinguish the elicitated combinations of focus condition—from left to right: narrow focus, broad focus and unfocused given words—and grammatical function / position of the word—from top to bottom: position 1, 2 and 3, i.e. subject, verb and object. Within each panel, the first syllable vowel quantities for the particular subgroup are compared. The upper bar shows the average durations for words with long first syllable vowels (quantity 2), while the lower bar illustrates those for quantity 1 words. As related in section 2.2.1, the quantity of the second syllable is interrelated with a word's grammatical function and thus with its position in the sentence. Figure 2.11 therefore codes first and second syllable quantity together next to each bar. For example, 'Q2Q1' signifies words where the first syllable vowel is long, but the second syllable vowel is short, a combination that only appears in sentence-initial position (subjects).

Figure 2.11 illustrates that all three experimental factors—focus condition, quantity and position—had an influence on overall word duration. Accordingly, the linear mixed-effects model indicated a main effect for each factor (see Table 2.9). Compared to broad focus, word duration was longer in narrow focus and shorter for unfocused given words. To reduce complexity, the two types of focus (contrastive and information focus) are not distinguished in Figure 2.11, but the model indicates that the effects were found for both types.

As for quantity, words with long first syllable vowels had a significantly longer overall duration than those with quantity 1 vowels. The model in Table 2.9 only included first syllable vowel quantity as a factor, since second syllable vowel quantity was correlated with position. In the absence of statistical testing, note that the mean durations of quantity 2 vowels were also longer in second syllables, as illustrated in Figure 2.11.

Also, duration was longer in the third position, i.e. for the objects. Table 2.9 indicates that this effect was statistically significant. It is also visible in Figure 2.11. It should especially be pointed out that Q2Q2 words, appearing both in position 2 and in position 3, had a longer mean duration in sentence-final position. The generally longer duration in sentence-final position also interacted with focus condition, so that narrow focus lengthening was less strong in position 3. The shortening of unfocused sentence-final words was more intensive, at least when the sentence included a contrastive focus elsewhere (see the significant interaction between unfocused (CF) and position 3 in Table 2.9). By contrast, the durational lengthening effect of narrow focus was larger for verbs, while the shortening in unfocused condition was smaller. Thus in position 2, broad focus (the intercept) was more like unfocused and less like narrow focus than for the other grammatical functions / positions.

Additionally, position interacted with quantity: The effect of quantity was larger in medial and final position than for the sentence-initial intercept of the model. Finally, the interactions near the bottom of Table 2.9 suggest that the durational lengthening effect of narrow focus was larger in the overall longer quantity 2 words.

On a whole, this section has not only shown that word duration is affected by all three experimental factors—focus condition, quantity and position—but also that the effects of these factors interact in several ways.

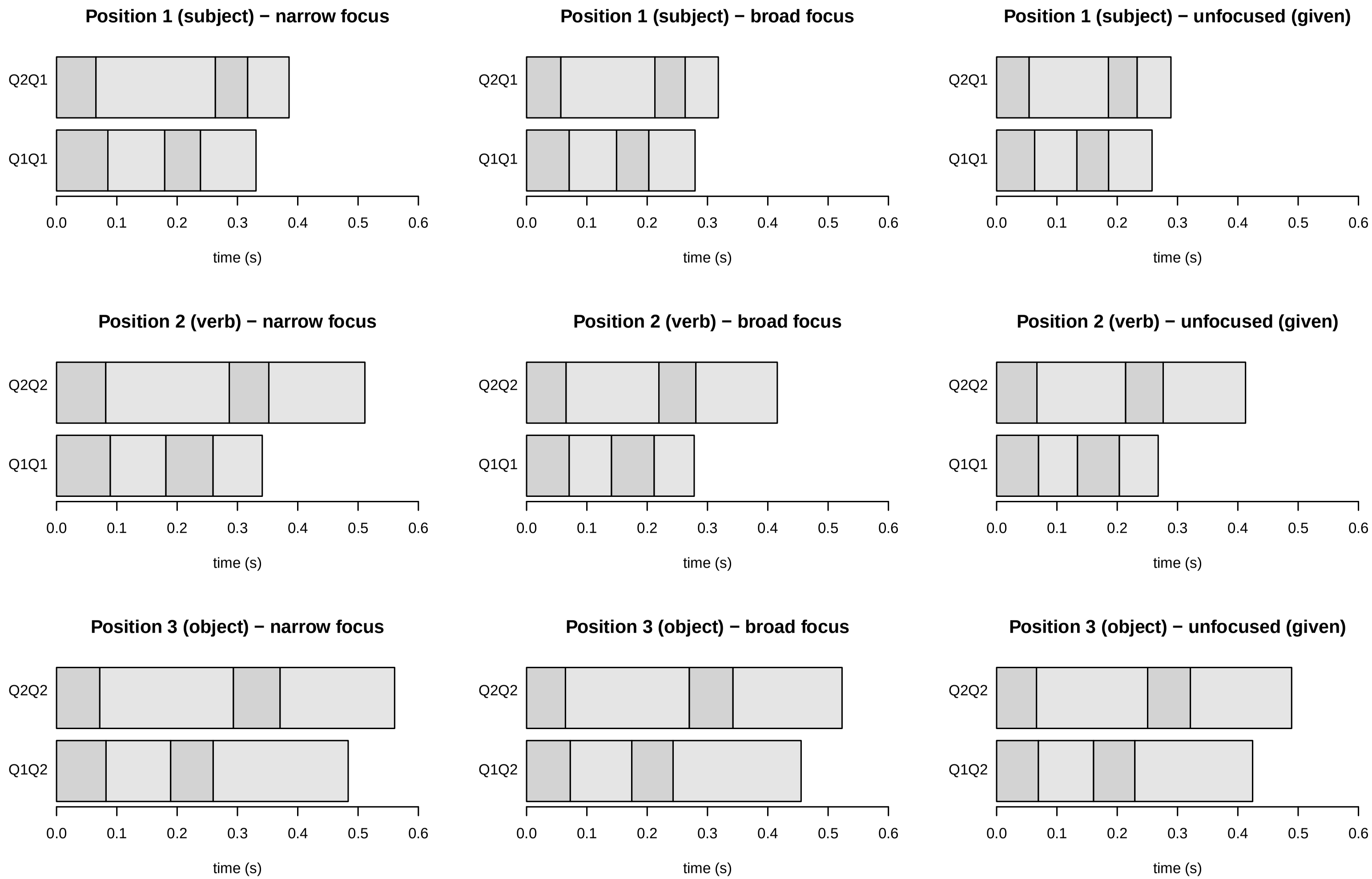

Figure 2.11: Average segment durations for all experimental conditions.

Table 2.9: Coefficients of the best linear mixed-effects model of word duration (in ms) with random effects of item and of focus condition, (first syllable) quantity, position and trial by subject. CF indicates contrastive focus, IF information focus on the narrow focus constituent. The factor trial was centred to its median.

	Estimate	Std. Error	t value
(Intercept)	280.1203	9.1899	30.4815
Narrow focus (CF)	57.6798	10.5160	5.4849
Narrow focus (IF)	44.3635	9.3260	4.7570
Unfocused (CF)	−19.0279	5.4776	−3.4738
Unfocused (IF)	−27.4645	5.5141	−4.9807
Position 2	−7.8519	9.9243	−0.7912
Position 3	175.5715	11.8242	14.8485
Quantity 2	36.9113	8.9921	4.1048
Trial	−0.0108	0.0549	−0.1958
Narrow focus (CF) : position 2	18.1277	6.6741	2.7161
Narrow focus (IF) : position 2	20.6674	6.6653	3.1007
Unfocused (CF) : position 2	18.3044	5.7735	3.1704
Unfocused (IF) : position 2	19.4944	5.7675	3.3801
Narrow focus (CF) : position 3	−23.5388	6.6824	−3.5225
Narrow focus (IF) : position 3	−27.5471	6.6578	−4.1376
Unfocused (CF) : position 3	−11.9480	5.7723	−2.0699
Unfocused (IF) : position 3	−2.7138	5.7690	−0.4704
Position 2 : quantity 2	111.8270	10.7180	10.4336
Position 3 : quantity 2	29.0769	10.7180	2.7129
Narrow focus (CF) : quantity 2	18.4994	5.4775	3.3774
Narrow focus (IF) : quantity 2	15.0629	5.4709	2.7532
Unfocused (CF) : quantity 2	−5.4946	4.7539	−1.1558
Unfocused (IF) : quantity 2	−2.5694	4.7506	−0.5409

2.4.4 Alignment

This section discusses the alignment of the three measured f_0 turning points—L1, H and L2—in two ways. First, it evaluates segmental alignment, showing how many of the respective minima and maxima occurred on different segments. Second, it shows alignment with more temporal precision, illustrating absolute mean alignment relative to the beginning of the word per condition. Statistical analysis of alignment as the temporal distance from several different anchor points is postponed until the next chapter (see especially section 3.2, focusing on peak alignment; alignment of minima is evaluated statistically for another data set in Chapter 4). The main hypothesis in section 2.3.4 was that there is no phonological contrast in peak alignment emerging from the experimental variations. Additionally, an analysis in terms of phrase tones was hypothesised to be superior over accounts employing uniform L+H* or LHL accents.

For segmental alignment, Appendix D contains detailed results for the number and percentage of measured f_0 turning points on the words' first syllable onset consonant (C1), first syllable vowel nucleus (V1), second syllable onset consonant (C2) and second syllable vowel nucleus (V2). For example, when the f_0 maximum of a word was detected during the first syllable vowel nucleus, as for the objects in Figures 2.3 and 2.10, this was counted as a case where the measurement point H was aligned with V1 (nuclei consisted of a long or short vowel or a diphthong as defined in section 2.2.1). Overall, these results indicated that L1 was most frequently realised on the onset consonant C1, L2 appeared most often on the second syllable vowel V2 and H was most often measured during the first syllable vowel V1. However, the experimental conditions influenced segmental alignment, modulating the predominance of the most common alignment for the three measurement points. Again, the two types of narrow focus generally showed little difference in alignment.

For L1, the alignment to the onset consonant was most consistently realised in the narrow focus conditions (overall in 95% of the cases, cf. Figure 2.12). Within this information structural condition, more C1-alignment was found in position 3 and position 2 than on the words in first position (100%, 97% and 85%, respectively). In comparison to narrow focus, broad focus words generally showed less localisations on the first consonant, with an overall drop of about 10% in position 3, 15% in position 1 and almost 30% in position 2. Only 62% C1-realisations emerged for unfocused given words, with a large variation in alignment especially in second and third position. No systematic differences between the quantities occurred regarding the alignment of L1, but Figure 2.12 shows a much lower percentage of C1-alignments for quantity 2 than for quantity 1 in position 1 under broad focus. While in this condition, only 58% of L1 turning points were measured on C1 for quantity 2, the remaining 42% appeared on the next segment, the

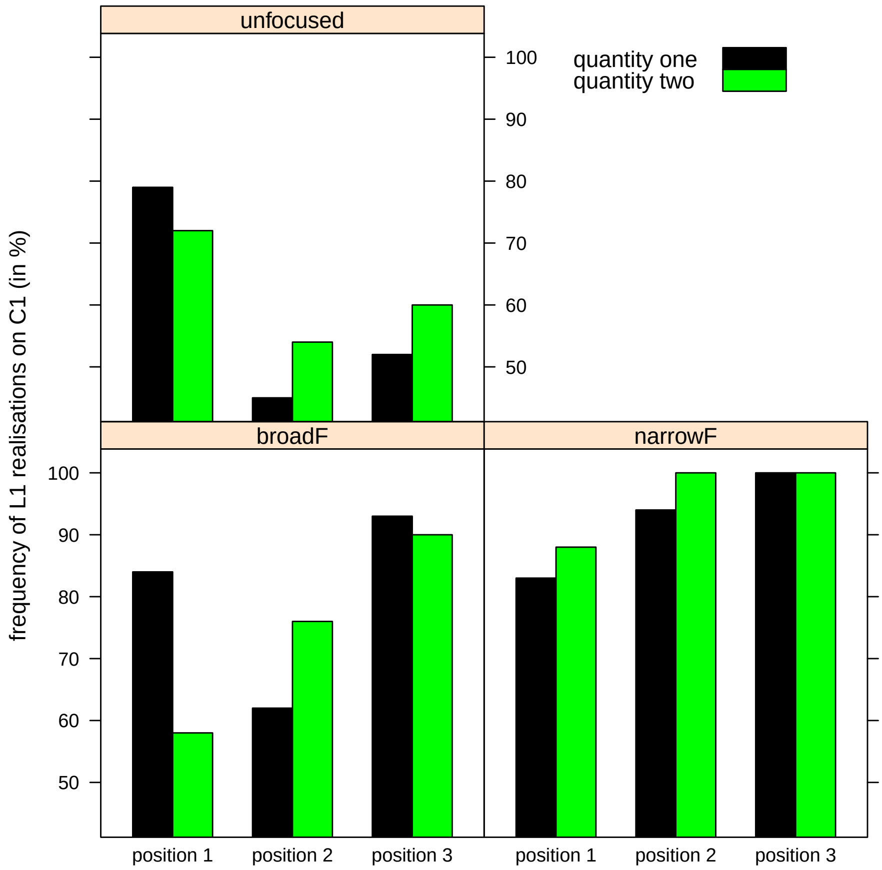

Figure 2.12: Percentage of L1 measurement points detected during the word's first consonant (C1) for different experimental conditions.

first syllable vowel nucleus V1. In quantity 1, where first vowel duration was shorter, 84% of L1 points were measured on C1 and 16% on V1.

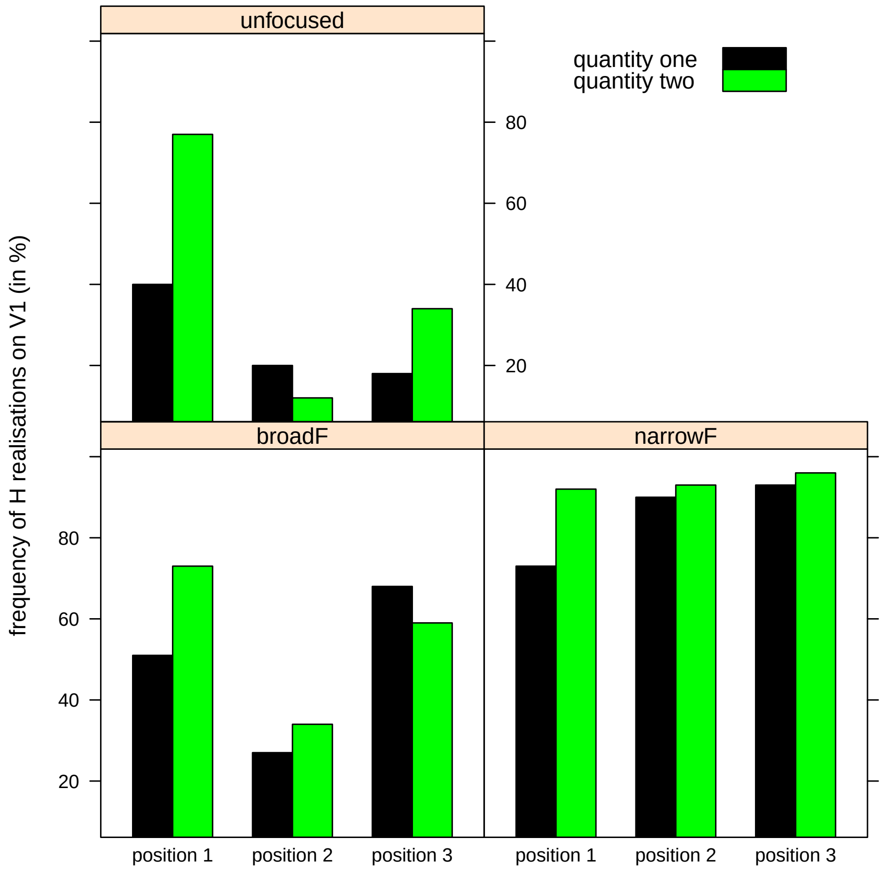

Figure 2.13: Percentage of H measurement points detected during the word's first vowel (V1) for different experimental conditions.

For the f_0 maximum H, the most common alignment was to the first syllable nucleus V1. The highest percentage of V1-realisations was found in narrow focus (altogether 90%), while peak alignment on unfocused words was least unified, with only 36% alignment to V1. In broad focus, overall 53% of pitch maxima appeared on the first vowel nucleus.

Figure 2.13 also shows a difference between positions, interacting with focus condition. In sentence-medial position, the variation in peak location was largest. In narrow focus, almost all f_0 maxima of verbs were measured during the first syllable vowel. For broad focus, only 31% of H points in position 2 appeared on V1, while 24% were found on first syllable onsets and 38% on second syllable vowels (see the illustrations of early and late maxima on verbs for example in Figures 2.2, 2.4 and 2.5). For unfocused verbs, f_0 maxima were located on C1 and V2 about equally frequently (40% and 38%), while only 15% of peaks were found on V1. Words in position 3 had peaks during V1 in 94% and 63% of the cases in narrow focus and broad focus, respectively, but more variation appeared in unfocused condition. Only 26% of H points were found on V1, while for between a third and half of the unfocused sentence-final words, f_0 maxima were located on the second syllable nucleus V2. Upon closer inspection, these late maxima often seemed perceptually different from the late maxima in other positions and rather appeared to be sentence-final rises, although this was not always clear (see Ogden and Routarinne, 2005, on sentence-final rises as a continuation signal in Finnish). For 15% of the utterances (137), I was sufficiently sure that a final rise was present. Figure 2.14 provides an example; also confer the discussion in Chapter 3.

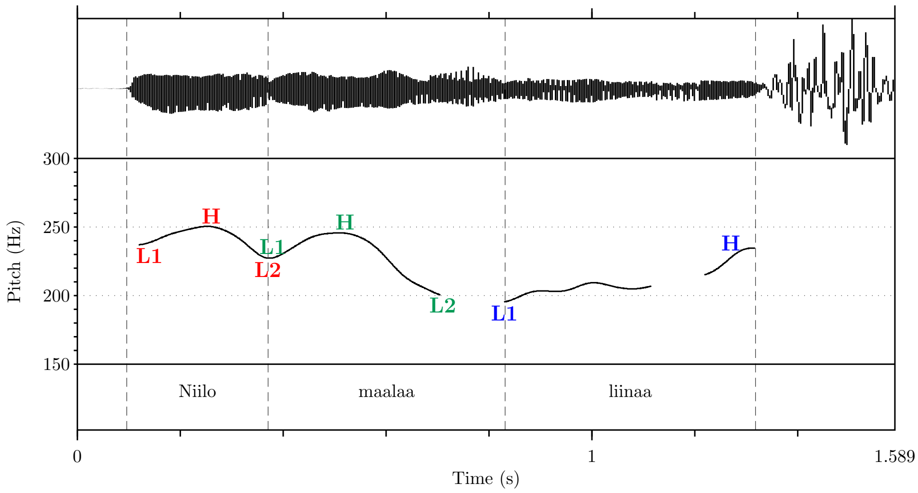

Figure 2.14: Contour with sentence-final rise. Sentence *Niilo maalaa liinaa* 'Niilo paints a cloth', with narrow information focus on the verb.

Finally, differences between the quantities were generally not large or systematic, with one exception. In sentence-initial position, the percentage of H measurements on V1 was consistently higher in quantity 2 than in quantity 1 (overall 81% vs. 51%), whereas quantity 1 words showed more peaks on the following consonant C2 than quantity 2 words (21% vs. 1%). Nevertheless, alignment to the first syllable vowel was still the overall most frequent finding for words in position 1, regardless of quantity (or information structure).

The second low f_0 measurement point L2 was overall most often measured on the second syllable vocalic nucleus. The predominance of V2 alignments was clearest in position 1 (overall 86%), while there was more variation in

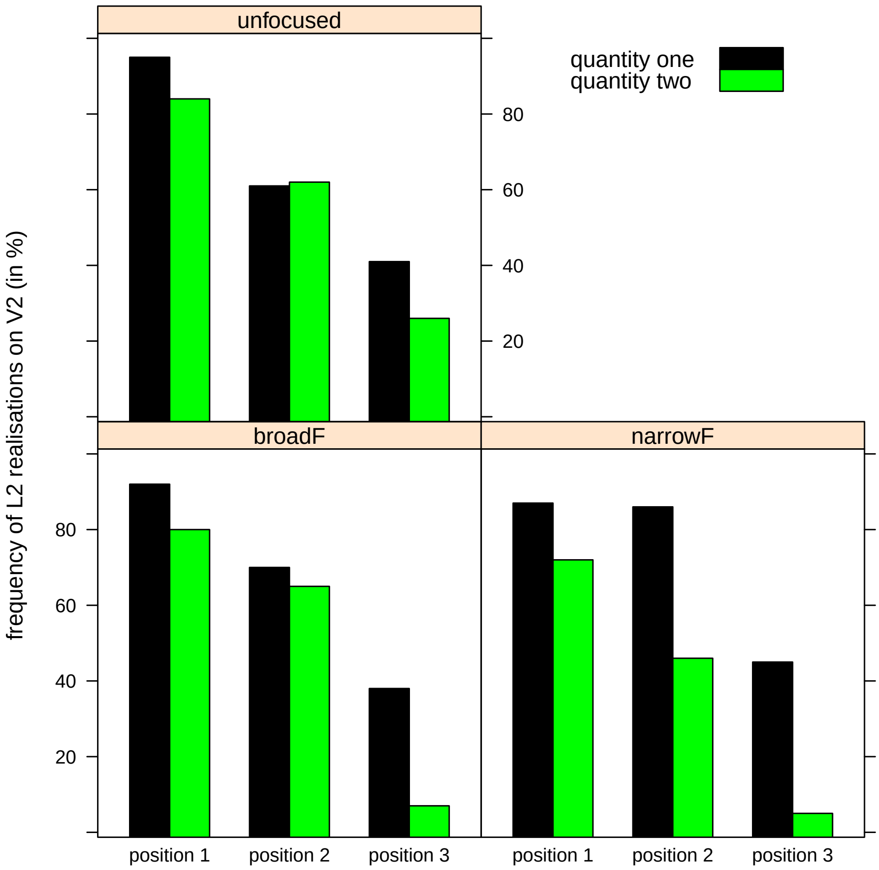

Figure 2.15: Percentage of L2 measurement points detected during the word's second vowel (V2) for different experimental conditions.

position 2 (64% V2-alignment). In position 3, L2 was often marked on earlier segments, with only 27% V2-alignment overall. Figure 2.15 illustrates that the difference between the positions appeared for both quantities and in all focus conditions. Additionally, there was an effect of focus condition modulated by position: Alignment to V2 was found for overall 71% of unfocused words and for 57% and 56% in broad and narrow focus, respectively. However, this effect only appeared in positions 1 and 3, while L2 of sentence-medial verbs was measured on V2 on 68%, 65% and 61% of broad focus, narrow focus and unfocused realisations, respectively. Finally, there were also differences between the quantities: Figure 2.15 shows a higher percentage of V2-alignments for quantity 1 in all but one condition (overall, L1 was measured on V2 for 72% of quantity 1 and 53% of quantity 2 words). Conversely, 31% of L2 appeared on the first vocalic nucleus V1 in quantity 2, i.e. when V1 consisted of a long vowel or diphthong (vs. 9% and in quantity 1).

A more detailed picture of alignment emerges from Figure 2.16, which combines a visualisation of segment durations as already familiar from Figure 2.11 with an illustration of mean times of f_0 measurement points. Thereby, black diamonds symbolise the mean times of the pitch maxima H and the grey squares display the mean times of the minima before and after, i.e. L1 and L2, respectively. Again, darker bars indicate consonant and lighter ones vowel durations.

The figure suggests a distinction between two kinds of alignment patterns occurring in different conditions, marked with different colour schemes. Interestingly, one pattern is found for broad focus objects (position 3) in addition to narrowly focused words in all three positions, while the other pattern appears in all other cases. Therefore, the panels with green segment duration bars will be referred to as showing the nuclear conditions, whereas panels with blue colours display alignment for non-nuclear conditions.

In terms of the low turning points, the two alignment patterns were rather similar. Overall, the first low f_0 points L1 were located close to the beginning of the word, while the mean time for the second low L2 was more or less close to the end of the word. Within each position in the sentence, the mean times of L1 points in nuclear conditions were a bit earlier than those of the non-nuclear ones. The L2 points were also measured a bit earlier for nuclear than for non-nuclear conditions and in general earlier in the third position objects than for subjects and verbs.

The most noticeable difference between the two patterns was the mean alignment of the H points. Peaks were generally realised earlier for nuclear conditions than for non-nuclear conditions. Also, while mean peak timing was very similar for both quantities in nuclear conditions, quantity 2 words had later mean H alignment than quantity 1 words in almost all non-nuclear conditions. Another way of formulating the difference is to relate peaks to two different landmarks, the beginning and the end of the first syllable

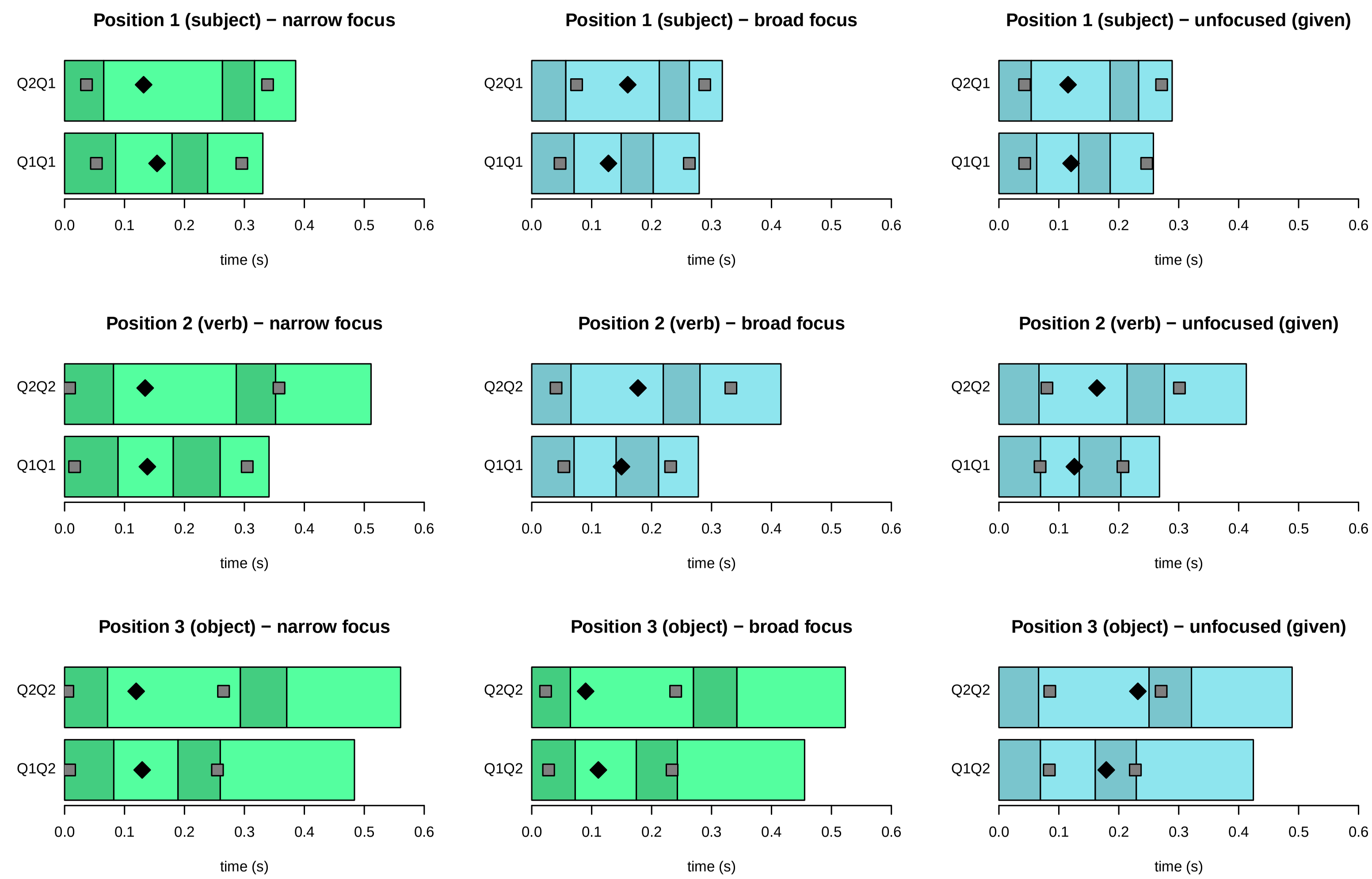

Figure 2.16: Average segment durations and times of f_0 turning points for all experimental conditions.

nucleus. Pitch maxima in non-nuclear conditions were on average closer to the end of the first syllable nucleus than those in nuclear conditions. At the same time, the differences between the quantities appearing in the non-nuclear conditions were smaller relative to the end of the nucleus than with respect to the beginning of the nucleus. However, note that Figure 2.16 only displays mean timings, not the whole distribution of measured values. Peak timing showed more variation for broad focus and unfocused words, whereas values from narrow focus words exhibited a narrower distribution. One-sided paired by-subject and by-item t-tests over standard deviations (SD) indicated that the difference was significant, both for peak alignment measured relative to the beginning of the first syllable nucleus ($t(16) = 9.85, p < .001$ and $t(23) = 6.26, p < .001$) and for the distance relative to the end of the first syllable nucleus ($t(16) = 11.23, p < .001$ and $t(23) = 6.40, p < .001$). The difference was also was also significant for alignment relative to the beginning of the word: ($t(16) = 9.36, p < .001$ and $t(23) = 5.94, p < .001$; see section 3.2 for further statistical evaluation of peak alignment in the present data set).

2.4.5 Voice quality

Voice quality was evaluated for all syllables in the data in a binary way, i.e. categorising it as either modal or non-modal. Section 2.3.5 predicted that non-modal voice quality would be more frequent towards the end of the sentence, i.e. in position 3, than for the sentence-initial and sentence-medial constituents.

Figure 2.17 illustrates the percentage of syllables that were (partly or completely) realised with non-modal voice quality for each position in the sentence. For example, the bars labelled 'Obj_1' refer to the first syllable of the objects (position 3). The different basic focus conditions are indicated by the colours of the bars. For the two syllables of the second-position words, an additional distinction between pre- and post-focally given unfocused words is made.

A higher percentage of non-modal realisations at the end of the sentence can clearly be observed here. Interestingly, however, information structure crucially modified the effect of position. In broad focus, the first two words were mainly spoken with modal voice, while the sentence-final position exhibited a large amount of non-modal realisations. In other information structures, increased percentages of non-modal realisations also appeared in non-final position. The pattern can be generalised in the following way: An increased percentage of non-modal realisations was found on the second syllable of a narrowly focused words and all following syllables (see the schematic representation in Figure 2.18).

In line with this generalisation, separate linear mixed-effects models were fit for the first syllables of the three words (designated as 'Subj_1', 'Verb_1'

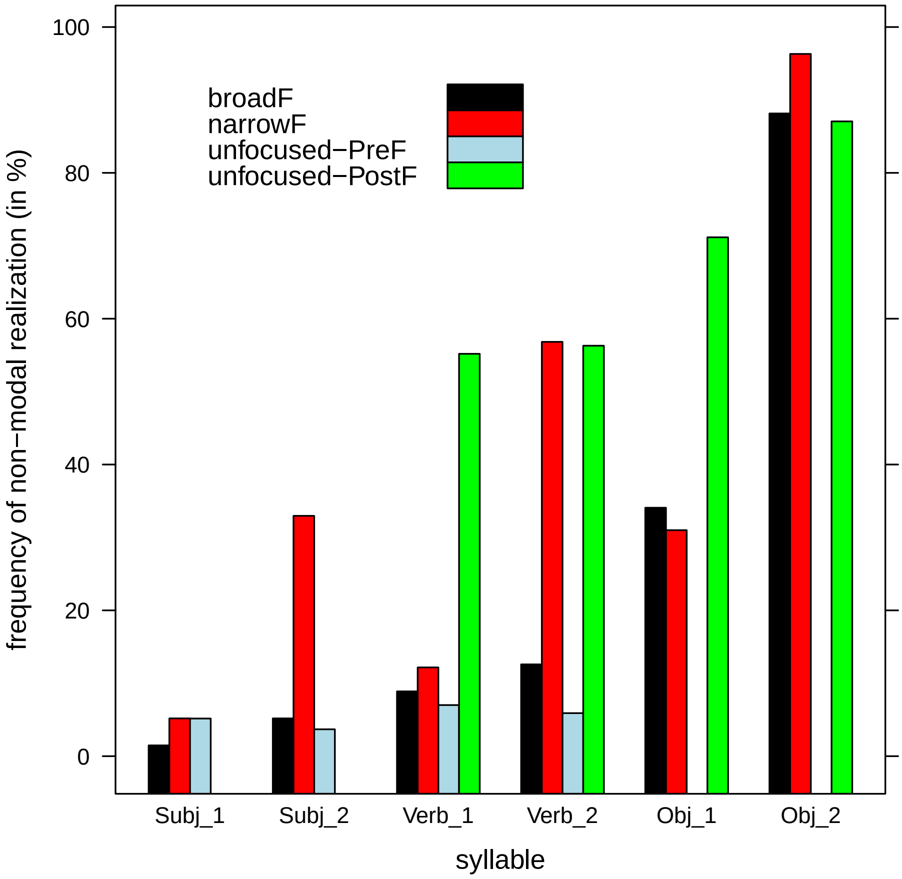

Figure 2.17: Frequency of usage of non-modal voice quality (in %) for syllables in all positions as a function of focus condition.

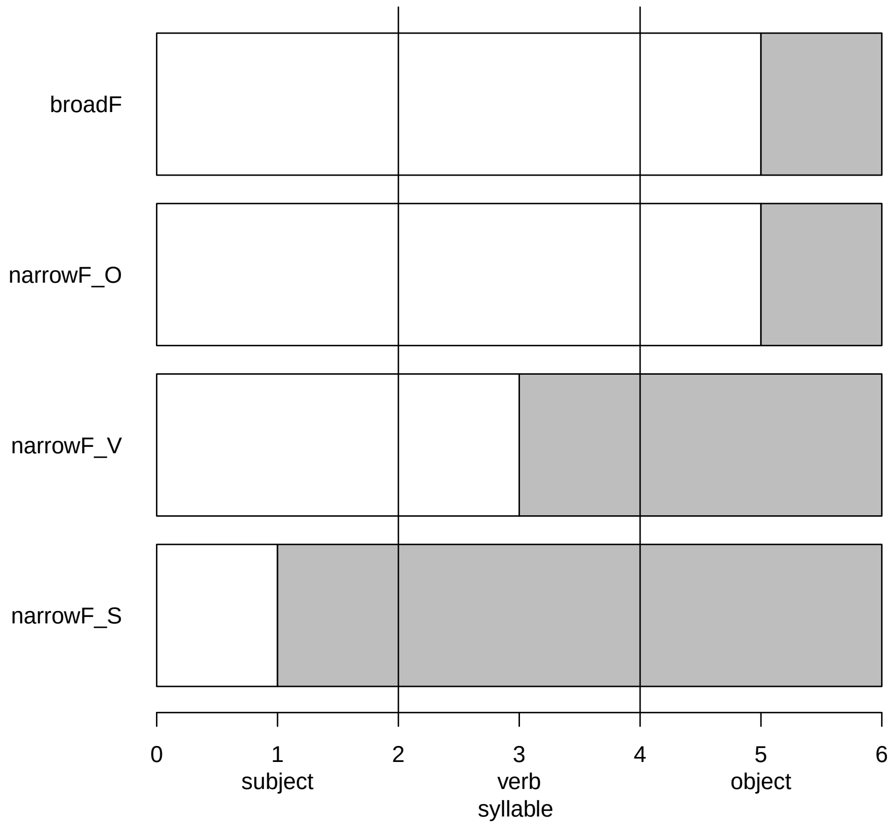

Figure 2.18: Schematic representation of the usage of non-modal voice quality (grey areas) in different information structures.

Table 2.10: Coefficients of the best binomial linear mixed-effects model of voice quality for the first syllables of words, with random effects of item and of position by subject; CF = contrastive focus, IF = information focus.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−3.4254	0.5011	−6.8358	0.0000
Position 2	0.1932	0.5316	0.3634	0.7163
Position 3	0.8612	0.5835	1.4759	0.1400
Quantity 2	−0.0407	0.5389	−0.0756	0.9397
Narrow CF	−0.8703	0.4645	−1.8736	0.0610
Narrow IF	−1.9148	0.6616	−2.8942	0.0038
Unfocused post-focal (CF)	3.8420	0.3458	11.1095	0.0000
Unfocused post-focal (IF)	3.0194	0.3357	8.9934	0.0000
Unfocused pre-focal (CF)	−0.3652	0.4687	−0.7791	0.4359
Unfocused pre-focal (IF	0.2994	0.4121	0.7265	0.4675
Position 2 : quantity 2	0.8588	0.5396	1.5917	0.1115
Position 3 : quantity 2	2.2432	0.5687	3.9441	0.0001
Quantity 2 : narrow CF	1.6004	0.5469	2.9263	0.0034
Quantity 2 : narrow IF	2.5495	0.7219	3.5314	0.0004
Quantity 2 : post-focal (CF)	−1.1459	0.4478	−2.5591	0.0105
Quantity 2 : post-focal (IF)	−1.1470	0.4333	−2.6469	0.0081
Quantity 2 : pre-focal (CF)	−0.3874	0.6372	−0.6080	0.5432
Quantity 2 : pre-focal (IF)	−0.6189	0.5665	−1.0924	0.2747

and ‘Obj_1’ in Figure 2.17, see Table 2.10) and for the second syllables (‘Subj_2’, ‘Verb_2’ and ‘Obj_2’, see Table 2.11). Since the dependent variable for these models is binary (modal vs. non-modal), binomial models were employed. These models estimated the logit transformed probability of a non-modal realisation in different experimental conditions (Baayen, 2008, pp. 214–215). Positive estimates indicate more, negative ones less realisations with (partly) non-modal quality. These models distinguished between pre-focal and post-focal given words. As a result, it was not possible to test for interactions between the factors focus condition and position. However, an ANOVA comparison to the respective best models with free interactions between all three experimental variables indicated that the difference was insignificant, while the models presented here even provided a slightly better fit. As observed in previous sections, effects were the same for the two focus types (information focus and contrastive focus) in both models.

For the first syllables, the amount of non-modal realisations was indeed only increased on post-focal given words, but not in pre-focal unfocused condition, whereas words in narrow focus were realised with non-modal voice

Table 2.11: Coefficients of the best binomial linear mixed-effects model of voice quality for the second syllables of words, with random effects of item and of position by subject; CF = contrastive focus, IF = information focus.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−3.1400	0.5248	−5.9829	0.0000
Position 2	1.1551	0.3156	3.6601	0.0003
Position 3	4.9834	0.6987	7.1327	0.0000
Quantity 2	−0.1531	0.4186	−0.3657	0.7146
Narrow CF	2.0595	0.2956	6.9671	0.0000
Narrow IF	1.5038	0.2903	5.1809	0.0000
Unfocused post-focal (CF)	1.9671	0.3195	6.1568	0.0000
Unfocused post-focal (IF)	1.4468	0.3052	4.7401	0.0000
Unfocused pre-focal (CF)	−1.6526	0.4884	−3.3833	0.0007
Unfocused pre-focal (IF)	−0.7905	0.3802	−2.0788	0.0376
Position 2 : quantity 2	0.6893	0.3881	1.7760	0.0757
Position 3 : quantity 2	−0.3216	0.4500	−0.7146	0.4748
Quantity 2 : narrow CF	0.1706	0.4124	0.4136	0.6791
Quantity 2 : narrow IF	0.9418	0.4128	2.2816	0.0225
Quantity 2 : post-focal (CF)	−0.5318	0.4286	−1.2407	0.2147
Quantity 2 : post-focal (IF	−0.2725	0.4133	−0.6594	0.5096
Quantity 2 : pre-focal (CF)	−0.0894	0.6691	−0.1336	0.8938
Quantity 2 : pre-focal (IF)	−0.5484	0.5560	−0.9863	0.3240

significantly less often (see Table 2.10). In contrast, second syllables exhibited more non-modal realisations both in post-focal condition and in narrow focus, while pre-focal givenness decreased the amount of non-modal realisations (see Table 2.11). This supports the conclusion that the use of creaky or breathy voice quality marking post-focal material already started on the second syllable of the narrow focus constituent itself. In Figure 2.17, this can most clearly be seen for the subject, where a noteworthy proportion of non-modal realisations of second syllables only appeared in narrow focus condition (see the red bar marked 'Subj_2').

Also the factor position affected realisations of first and second syllables differently. Non-modal realisations of a word's second syllable were overall more frequent later in the sentence (see Table 2.11). As an example, compare the percentages of non-modal voice in narrow focus in Figure 2.17, which show a gradual increase. A similar tendency can be discerned by comparing the bars for first syllables within most focus conditions, but the factor position did not reach significance as a main effect for first syllables. However, Table 2.10 indicates a significant interaction between position and quantity, with the first syllables of quantity 2 objects being realised as non-modal more often.

As an illustration, consider Figure 2.19, which is similar to Figure 2.17 but additionally distinguishes between words with different first syllable vowel quantities. In most of the cases, the percentage of non-modal realisations was virtually identical for the two vowel quantities. However, there were a few exceptions. First, a difference can be seen for the first syllable of the sentence-final words ('Obj_1') in broad focus and narrow focus (see the positive interaction position 3 : quantity 2 in Table 2.10). In the unfocused condition, the same distinction is much smaller, which might be due to the fact that the percentage of non-modal realisations is so high on a whole that there is not much room for realising a difference. Together with the absence of a difference in post-focal verbs, this probably accounts for the negative interaction between the post-focal conditions and quantity 2 in Table 2.10. Second, both syllables of the verb show a larger amount of non-modal realisations for quantity 2 in narrow focus. However, the increase did not reach significance in Table 2.10 and only resulted in a marginal interaction in Table 2.11.[8] A final interaction in Table 2.11 indicates that non-modal phonation on the second syllable of a word in narrow (information) focus was even more frequent in quantity 2.

2.4.6 Occurrence of pauses

In the three-word sentences of the current experiment, there are two possible positions for pauses in each utterance, one between the subject and the

[8] Note, though, that removing the interaction term resulted in a significantly worse model fit as assessed by an ANOVA comparison.

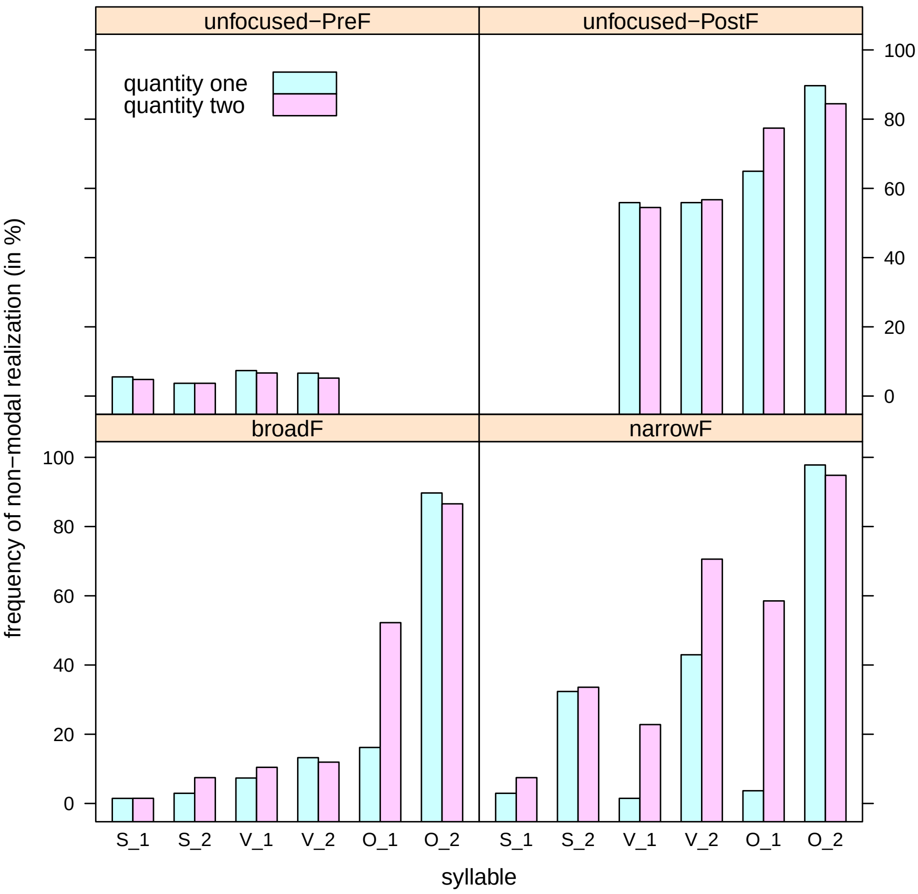

Figure 2.19: Frequency of usage of non-modal voice quality (in %) for syllables in all positions as a function of quantity by focus condition.

verb and one between the verb and the object. Section 2.3.6 tentatively predicted that the occurrence of pauses would vary by information structure and position in the sentence, but not be influenced by quantity.

Altogether, 111 pauses occurred in the 947 sentences of the present data set. Their distribution across the different information structures appears in Figure 2.20. Between subject and verbs, the number of pauses was highest with narrow focus on the subject—pauses occurred in 25% and 20% of the realisations for contrastive and information focus on the subject, respectively. Pauses between the verb and the object were most frequent when the verb was in narrow focus—they were found 13% and 7% of the cases in contrastive focus and information focus, respectively.

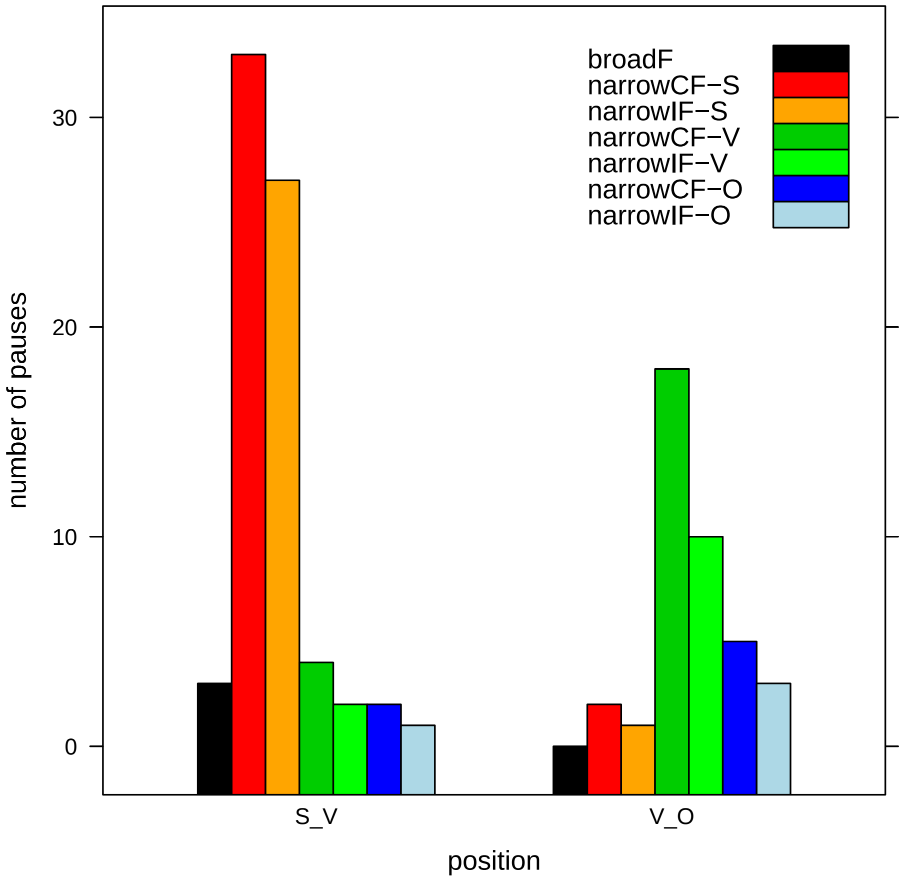

Figure 2.20: Number of pauses between subject and verb (S_V) and between verb and object (V_O) for different focus conditions.

The number of pauses was thus higher after a word when it was the narrow focus constituent. This was significant according to the model in

Table 2.12: Coefficients of the best binomial linear mixed-effects model of pause occurrence after subjects and verbs (position 1 and 2), random effects of item and of focus condition by subject. Positive estimates indicate a larger number of pauses.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−6.0108	1.2241	−4.9105	0.0000
Narrow focus (CF)	3.4497	1.5084	2.2870	0.0222
Narrow focus (IF)	2.9994	1.5067	1.9906	0.0465
Unfocused (CF)	0.1065	1.3469	0.0791	0.9370
Unfocused (IF)	0.3921	1.4234	0.2754	0.7830
Position 2	−15.1332	1188.0846	−0.0127	0.9898
Narrow focus (CF) : position 2	13.7299	1188.0846	0.0116	0.9908
Narrow focus (IF) : position 2	13.4329	1188.0846	0.0113	0.9910
Unfocused (CF) : position 2	15.3508	1188.0847	0.0129	0.9897
Unfocused (IF) : position 2	15.4651	1188.0848	0.0130	0.9896

Table 2.12. Figure 2.20 suggests that the effect was slightly stronger for contrastive than for information focus and that overall less pauses appeared after verbs than after subjects, but these differences did not prove significant. There is, however, some indication that the difference between subjects and verbs was relevant, since removing the factor position and / or the interaction resulted in a significantly worse model. Also, a less optimal alternative model with a simple random effects structure showed a significant main effect of position.

In contrast, adding vowel quantity as a factor did not improve the model, suggesting that it did not affect the occurrence of pauses.

2.5 Discussion and summary

This chapter has investigated the effects of vowel quantity, focus condition and position / grammatical function on six prosodic parameters: identifiability of maximally three f_0 points (L1, H and L2) on each word, the pitch scaling of these points, word duration, the alignment of the f_0 measurement points, voice quality and the occurrence of pauses. The following subsections will summarise the effects of the three experimental variables in turn. Each subsection will then discuss the relevance of the respective findings to the chapter's research questions, repeated below. The discussions will evaluate how well the data support the hypotheses given in section 2.3 or their alternatives. Finally, a new alternative account—the nuclear accent hypothesis—will be presented in section 2.5.4.

Qu1 Can the data be accounted for by just one type of tonal specification on the phonological level or do they suggest a contrast between two or more accents?

H1 There is tonal uniformity at the phonological level.

A1 There are contrasting specifications (different types of accent).

Qu2 Is an analysis in terms of phrase tones or accents preferable for this data?

H2 Phrase tones account better for the data.

A2 Accents account better for the data.

Qu3 What is the shape and tonal specification of the contour(s)?

H3 Contours are specified as falling ($\mathrm{H_pL_p}$).

A3 Contours are specified as rising-falling (LHL) or rising (L+H*).

2.5.1 Effects of position

This section discusses the observed differences between sentence-initial subjects (position 1), sentence-medial verbs (position 2) and sentence-final objects (position 3; recall that due to the use of sentences with unmarked word order, position was correlated with grammatical function). These were found for all investigated parameters:

Tonal contours showed more variation in sentence-medial position and verbs carried a complete rise-fall contour of their own less often than constituents in other positions (see section 2.4.1).

Pitch scaling showed a general downtrend throughout the sentences. However, the drop in pitch was larger between the subject and the verb than between the verb and the object. At the end of sentence-final pitch falls, f_0 values converged across focus and quantity conditions (see section 2.4.2).

Word duration was longer in sentence-final position. Also, narrow focus lengthening had a larger effect and shortening of unfocused words a smaller effect in position 2 (see section 2.4.3).

Alignment of peaks to segments varied more in second and—in unfocused condition—third position than in position 1. In position 3, overall more second f_0 minima (L2) were detected relatively early in the word (see section 2.4.4).

Voice quality was more frequently non-modal later in the sentence, with over 80% non-modal realisations on the final syllable (see section 2.4.5).

Pauses between subjects and verbs were more frequent than pauses between verbs and objects (see section 2.4.6).

Most of these effects showed a special behaviour of the sentence-medial verbs—their realisation in broad focus was similar to that in unfocused condition. This was probably related to grammatical function, as it is reminiscent of the common notion of Finnish finite verbs being unaccented unless narrowly focused (see Välimaa-Blum, 1993). However, the distinction is interpreted here in terms of prosodic phrasing in accordance with **H2** above, (for the sake of simplicity, the present section concentrates on the p-phrase level). Thus, this section will argue that verbs constituted prosodic phrases of their own when narrowly focused, as shown in (13), but otherwise frequently formed p-phrases together with the following objects so that prosodic phrasing reflected syntactic phrasing, see (14). With regard to **Qu1**, the suggested account argues for an analysis in terms of a single tonal category, the phrase tones H_pL_p. Notice also that Välimaa-Blum's (1988) and Suomi et al.'s (2008) analyses in terms of accent use only one category, a uniform accent, likewise in agreement with **H1**.

(13) Typical distribution of p-phrases for narrow verb focus

$$\begin{array}{cccccc} H_p & L_p & H_p & L_p & H_p & L_p \\ | & | & | & | & | & | \\ \multicolumn{2}{c}{\text{Subject}} & \multicolumn{2}{c}{\text{Verb}} & \multicolumn{2}{c}{\text{Object}} \\ \multicolumn{6}{c}{(\text{Subject})_p\ (\text{Verb})_p\ (\text{Object})_p} \end{array}$$

(14) Typical distribution of p-phrases for other information structures

$$\begin{array}{ccccc} H_p & L_p & & H_p & L_p \\ | & | & & | & | \\ \multicolumn{2}{c}{\text{Subject}} & \text{Verb} & \multicolumn{2}{c}{\text{Object}} \\ \multicolumn{5}{c}{(\text{Subject})_p\ (\text{Verb Object})_p} \end{array}$$

As stated in **H3**, this thesis assumes that p-phrases are marked by H_pL_p tones. As for other phrase languages (Féry, 2010), the alignment of these tones is somewhat flexible. For instance, Figures 2.4 and 2.4 above showed examples of verb and object sharing a rising-falling pitch contour with the peak on the object in the former case and the peak on the verb in the latter case. While both are analysed as having one p-phrase spanning verb and object, the difference is understood to lie in the alignment of H_p to either the verb or the object, as shown in (15) and (16), respectively (see the annotated illustrations in Figures 2.21 and 2.22).

(15) Analysis for the example in Figure 2.21

H_p L_p H_p L_p

| | | |

Maini neuloi loimea

$(\text{Maini})_p$ $(\text{neuloi loimea})_p$

'Maini knitted a blanket.'

(16) Analysis for the example in Figure 2.22

H_p L_p H_p L_p

| | | |

Maini neuloi loimea

$(\text{Maini})_p$ $(\text{neuloi loimea})_p$

'Maini knitted a blanket.'

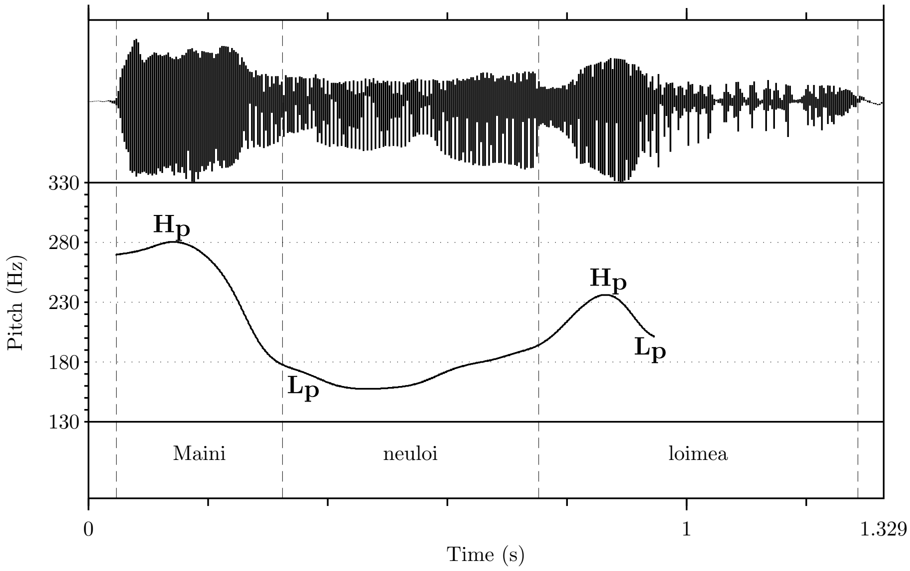

Figure 2.21: Annotation of verb and object as sharing an H_pL_p contour with the peak realised on the object. Sentence *Maini neuloi loimea* 'Maini knitted a blanket', broad focus (= Figure 2.4).

The special behaviour of the verbs—broad focus realisations resembling unfocused ones—emerged for most investigated parameters. This was the case for **duration**, with the durational lengthening effect of narrow focus being larger for verbs, while the shortening in unfocused condition was smaller. This pattern fits well with the phrasing account assuming that verbs formed p-phrases of their own in narrow focus, but were non-final in larger p-phrases in the other two conditions.

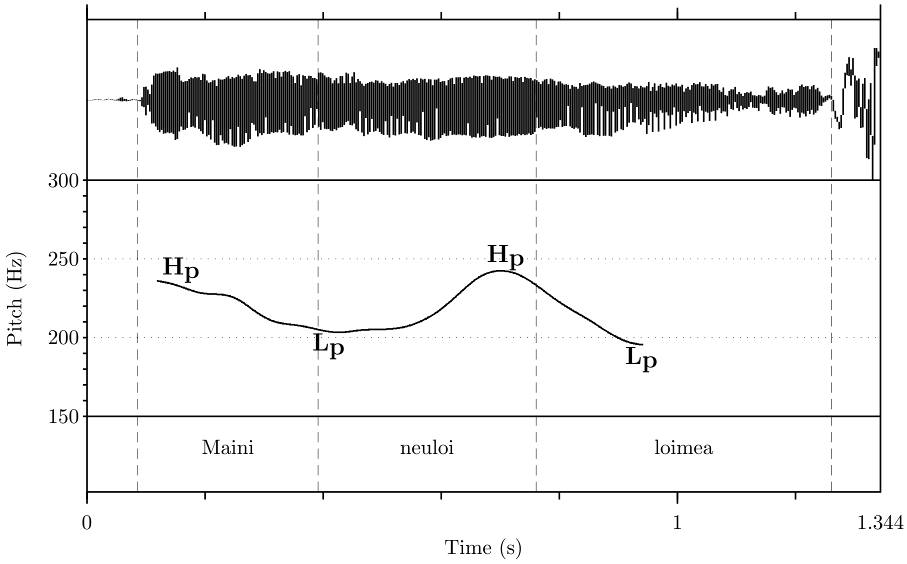

Figure 2.22: Annotation of verb and object as sharing an H_pL_p contour with the peak realised on the verb. Sentence *Maini neuloi loimea* ‘Maini knitted a blanket’, broad focus (= Figure 2.5).

The same distinction emerged for **pitch scaling**. As section 2.3.2 predicted, pitch values generally showed a downtrend throughout the sentence.[9] However, the lowering of the f_0 values for L1 and H relative to the sentence-initial subjects was not much larger for the sentence-final objects than for the verbs. Put differently, the step down from the first to the second position was generally larger than that from position 2 to position 3. In spite of this, the mean peak values of narrow focus constituents in Figure 2.6 form a continuously declining line, indicating again that verbs in broad focus and unfocused condition resembled each other. The difference is again straightforward in the present account: While f_0 maxima of narrow focus verbs corresponded to H_p tones, maxima of verbs in other focus conditions were often parts of interpolations between tonal targets on the neighbouring words.

Verbs were also special in showing the largest variation in terms of **tonal contour**, i.e. the identifiability of the measurement points L1, H and L2. The generally dominant combination LHL was least often found on verbs. This could be due to the fact that the lexical material only included plosives in verbs—plosives at the word onset made the realisation or measurement of L1 impossible for some verbs. However, there is evidence that this is not

[9]Whether this downtrend should be described as declination, as downstep or as a combination of both is left open here, since this question would require an empirical investigation (an example of such a study for Mexican Spanish is Prieto et al., 1996).

the whole picture. Depending on the location of narrow focus on either the preceding subject or the following object, unfocused verbs were found to display more HL or more LH contours. This shows that the pitch contour and thus the combination of measurable f_0 points was influenced by the contours of the neighbouring constituents. The subjective evaluation of tonal contours likewise indicated that the pitch contour of a word was most often determined by its neighbours in the second position—on verbs, pitch contours were judged as rises to the following word's peak more often than on subjects and as falls from the preceding word's peaks more often than on objects. Also the distribution of H point **alignments** on verbs in broad focus was more similar to the unfocused given condition than to narrow focus. In particular, the distribution of H locations across segments showed a larger variation for both broad focus and unfocused verbs than for verbs in narrow focus. Here, these variations are interpreted as due to variability in phrasing and flexibility of phrase tone locations even for a given phrasing, as illustrated in Figures 2.21 and 2.22. The differences in contour and alignment specifically of the verbs do however indicate that H_p tones were more frequently realised on the objects, as illustrated in (15).

Finally, **pauses** after verbs—marking prosodic phrase boundaries within the VP—occurred less frequently than pauses before verbs—marking prosodic phrase boundaries coinciding with syntactic phrase boundaries. This agrees with the tentative prediction that syntax would influence prosodic phrasing also within the short SVO sentences (see section 2.3.6).

Several of these findings could alternatively be explained in an accent-based account, as stated in **A2** above, assuming that broad focus and unfocused verbs are unaccented. This kind of approach might be appropriate for the results regarding tonal contours, pitch scaling and alignment, arguing that pitch on 'unaccented' verbs were interpolated and thus depended on the neighbouring words. However, it is not clear how one would account for the durational differences between 'accented' and 'unaccented' verbs, since accents are understood as tonal configurations in the AM-framework. Also the lower frequency of post-verbal pauses immediately follows from the proposed account of prosodic phrasing, but would require an additional stipulation with an accent-based approach. Moreover, while the quantitative data indicated that H_p tones were realised on the object more often, the crucial point is that a realisation of the peak on the verb is equally possible as an instantiation of the same category (recall that both Figure 2.21 and Figure 2.22 constitute perfectly appropriate realisations of broad focus). This kind of variability is typical for phrase tones, which may or may not move from a phrase boundary to a metrically more prominent location. Accounting for the same finding in terms of pitch accents, one would need to say that either the verb or—less often—the object can be realised as unaccented in broad focus.

Lastly, there were also effects of position that concerned the sentence-final constituents. **Word duration** was longer in sentence-final position, as hypothesised in section 2.3.3. This is understood as a finality marker here (see Nakai et al., 2009; Suomi et al., 2008, pp. 108–109, on final lengthening in Finnish). As another finality marker, a larger percentage of non-modal **voice quality** realisations were observed later in the sentence, confirming the hypothesis formulated in section 2.3.5 in line with previous descriptions (Iivonen, 1998; Nakai et al., 2009). Sentence-final non-modal voice quality is interpreted here as the reason for effects on **pitch scaling** and L2 **alignment**: That the f_0 values of final lows were extremely similar across conditions is explained by creaky or breathy voice quality cutting the (measurable) pitch falls short. For the same reason, L2 was measured earlier in position 3 than for sentence-initial or sentence-medial constituents.

2.5.2 Effects of focus condition

Focus condition influenced the prosodic realisation of the recorded sentences pertaining to all investigated parameters, as summarised below.

Tonal contours were more often complete rise-falls, resulting in measurable LHL points, on words in narrow focus than in broad focus, while unfocused words displayed less complete LHL contours. Unfocused words also showed the highest percentage of flat contours (see section 2.4.1).

Pitch scaling resulted in higher f_0 maxima and preceding minima L1 and lower following minima L2 on narrow focus words (see section 2.4.2).

Word duration was longer in narrow focus and shorter in unfocused condition compared to broad focus (see section 2.4.3).

Alignment of f_0 maxima varied less in narrow focus than for the other conditions (see section 2.4.4).

Voice quality was non-modal more often on the second syllable of a narrow focus word and all following syllables in the same sentence (see section 2.4.5).

Pauses appeared more often following words in narrow focus (see section 2.4.6).

This section suggests that the observed effects can be modelled as two strategies of focus marking, adjusting salience and phrasing. Figure 2.23 schematically illustrates this analysis, symbolising narrow focus constituents with red lines and unfocused ones with grey ones. As shown in the right panel, the effects of narrow focus are partly described as a requirement on phrasing here, namely that the narrow focus constituent be followed by an

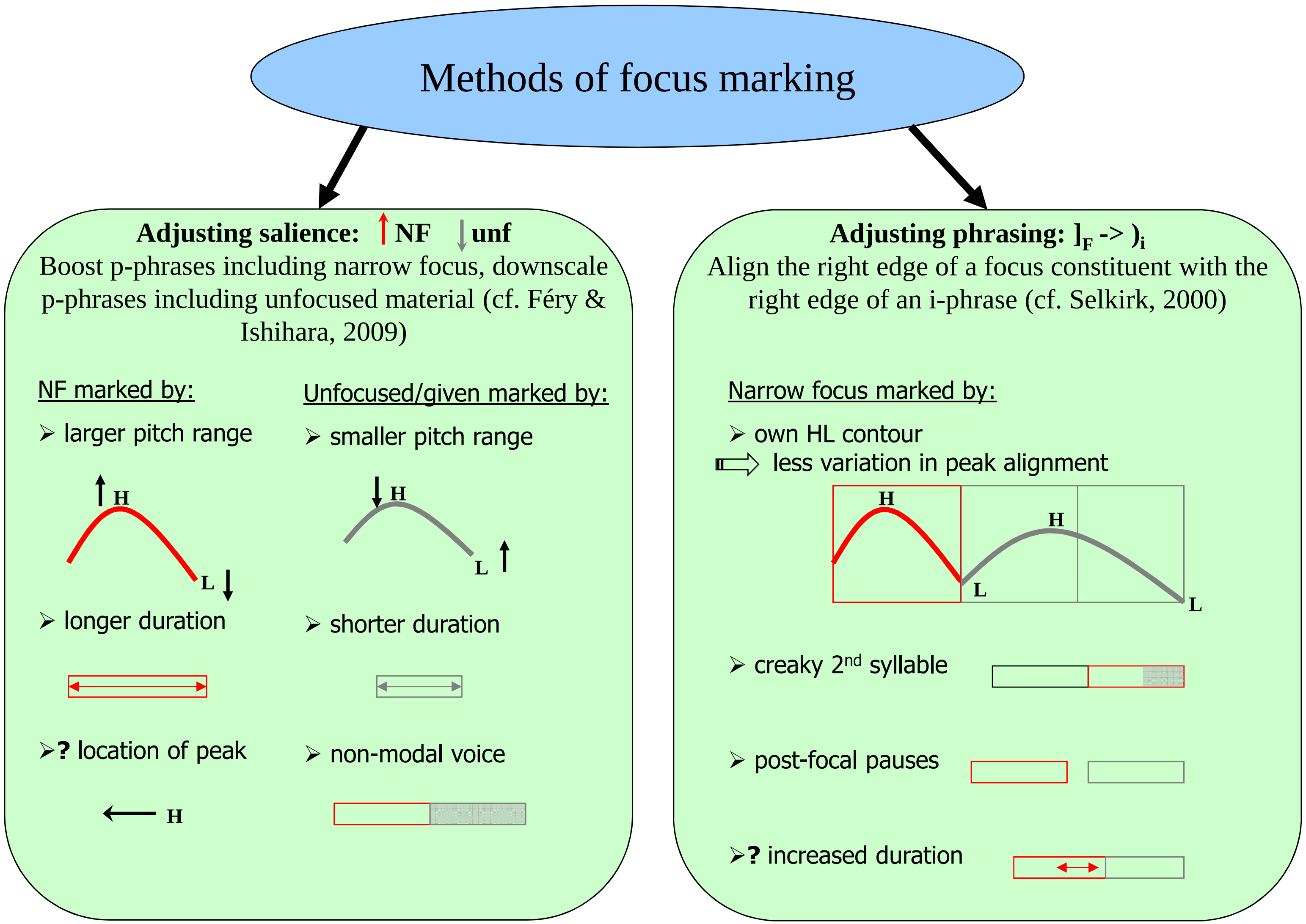

Figure 2.23: Schematic summary of observed prosodic correlates of focus condition.

i-phrase boundary. This could be modelled in an OT-framework employing a constraint like ALIGNFOC (Selkirk, 2000), but no attempt at a formalisation is made here. The suggested prototypical phrasing for the information structures elicitated in Production Experiment I appear in (17), where narrow focus constituents are marked by boldface. Note that the presence of an i-phrase boundary at the same time implies the presence of a lower-ranking p-phrase boundary. Also recall that this thesis assumes that broad focus or all-new sentences are focused as a whole (see section 1.2.2).

(17) a. Broad focus: $((\mathrm{S})_{\mathrm{p}}\ (\mathrm{VO})_{\mathrm{p}})_{\mathrm{i}}$

b. Narrow focus on the subject: $(((\mathbf{S})_{\mathrm{p}})_{\mathrm{i}}\ (\mathrm{VO})_{\mathrm{p}})_{\mathrm{i}}$

c. Narrow focus on the verb: $(((\mathrm{S})_{\mathrm{p}}\ (\mathbf{V})_{\mathrm{p}})_{\mathrm{i}}\ (\mathrm{O})_{\mathrm{p}})_{\mathrm{i}}$

d. Narrow focus on the object: $((\mathrm{S})_{\mathrm{p}}\ (\mathrm{V}\mathbf{O})_{\mathrm{p}})_{\mathrm{i}}$

The broad focus phrasing can be seen as a default that can for example be modified by narrow focus to align the right edge of the focus constituent with a i-phrase boundary. An additional phrase boundary thus arises for narrow verb focus (see (17c)), while a change is not required for object focus (see (17d)). Narrow focus on the subject inserts an additional i-phrase boundary, but does not affect the p-phrase level (see (17b)), since the subject NP and the VP already constitute different p-phrases in broad focus (see (17a)). The following discussion will return to this suggestion and review the evidence for it. At this point, notice the similarity between the narrow object focus in Figure 2.3 and the broad focus realisation in Figure 2.21. Figure 2.21 displays a downstep of the object's $\mathrm{H_p}$ target relative to the subject's $\mathrm{H_p}$, but otherwise, the contour does not differ from that in Figure 2.3.

As shown in Figure 2.23, narrow focus constituents are assumed to be right-edge aligned with i-phrases and, thereby, with p-phrases, regularly endowing them with $\mathrm{H_pL_p}$ tones of their own. This is taken to explain the finding of narrow focus words more frequently having complete rise-fall **tonal contours** of their own and thus measurable LHL points. At the same time, the realisation of an $\mathrm{H_p}$ tone on a narrow focus word resulted in earlier and more stable peak **alignment** compared to the location of f_0 maxima that were not realisations of phonological tones, but part of a stretch of interpolated pitch.

Section 2.3.5 stated that non-modal voice quality functions as a finality marker as ascertained by previous research (Iivonen, 1998; Nakai et al., 2009), but did not ascribe the marker to a particular level of prosodic phrasing. In this respect, is interesting to note that the frequency of non-modal **voice quality** realisations in the present data increased for later syllables rather than jump up abruptly on the sentence-final syllable(s) and, importantly, that it was additionally modified by information structure. As illustrated in Figure 2.23, second syllables of narrowly focused words were frequently realised by non-modal voice quality. The current data thus suggest that voice quality marks the finality of i-phrases.

These results indicate that post-focal **pauses** were used to demarcate the focus constituent and signal the right boundaries of i-phrases, though not obligatorily. The observed influence of focus condition fits the tentative hypothesis in section 2.3.6. Likewise as provisionally suggested, the difference between the positions point to a tendency to align the boundaries of prosodic phrases with those of syntactic phrases. One can thus say that a prosodic phrase boundary within the VP—i.e. between the verb and the subject—was disfavoured and appeared less often than a phrase boundary between the subject and the verb that kept the VP intact (or that it was marked less prominently, i.e. only by pitch range adjustments and voice quality, but not by a pause).

Figure 2.23 also suggests that the durational lengthening on narrow focus words probably was at least partially due to i-phrase finality. Although the distinction from salience-enhancing lengthening represented in the left panel is not perfectly clear, there is some indication for the presence of both. Sentence-final objects were arguably always i-phrase final (see (17)) and indeed showed a weaker effect of narrow focus lengthening in section 2.4.3. Also note that word durations were longer in both types of narrow focus, in contrast to Suomi et al.'s (2008) statement that only 'contrastive accents' receive lengthening.

Before turning to the left panel of Figure 2.23, it is important to note that the the phrasing indicated in (17) denotes only the assumed prototypical realisation of the respective information structures, which is subject to variation. For example the presence of two separate contours on verb and object in the broad focus realisation of Figure 2.10 indicates that it is possible to deviate from the default phrasing, represented for instance in Figure 2.21. Also, some remarks about the choice of referring to i-phrases and the role of recursion are in order. While the proposed prosodic hierarchy otherwise leaves the issue of recursion open (see Chapter 1), the effects of information structure on phrasing are accounted for with recursive i-phrases. This approach explains the findings better than a stipulation that p-phrases were aligned with focused constituents, especially regarding the different durational effects of narrow focus. Subjects, which were p-phrase final in narrow and broad focus according to the suggestion in (17), had longer durations in narrow focus only. While the extended duration could be ascribed either to a process increasing the salience of the narrow focus p-phrase or to the presence of a larger phrase boundary, the second account seems to be superior in the light of the durational effects of object focus. In sentence-final position, the effect of narrow focus was significantly reduced, which suits the assumption of lengthening as a finality phenomenon, since the objects are seen as i-phrase final in narrow focus as well as in broad focus. In contrast, if lengthening marked a p-phrase containing a narrow focus, this should in principle also affect i-phrase final p-phrases. Finally, the i-phrase could be seen as a domain of downstep, explaining the difference in pitch scaling of pre- and post-focal given material.

In addition to the adjustment of phrases, Figure 2.23 suggests that the relative salience of p-phrases containing narrow focus or unfocused material was adjusted in several ways (also see Féry and Ishihara, 2009). **Durational** lengthening in narrow focus was already mentioned above, additionally unfocused words were found to be shortened relative to broad focus realisations. Likewise, the findings for **pitch scaling** could be seen as adjustments attaining a larger pitch range for narrow focus (as predicted in section 2.3.2) and a smaller one for unfocused given words. Non-modal **voice quality** is also seen as having two functions, signalling the i-phrase finality of narrow focus words and making the post-focal p-phrase(s) less perceptually salient. Also the finding of a more consistent **alignment** of H tones to the (beginning of the) first syllable nucleus could be described as an attraction of the tone to a prominent location. It is however not quite clear that this assumption is needed here in addition to the suggested changes in phrasing.

To come back to the main research questions, the suggested analysis thus exclusively refers to prosodic phrases instead of accents, in accordance with **H2**. Regarding **Qu1**, it can be observed that the data support the general assumption of a single tonal contour. In particular, it is not possible to discern a contrast between—for example—a pattern A associated with a focus condition X and a pattern B dominating realisations of a focus condition Y. If, say, narrow focus were usually realised with a falling accent and words in broad focus carried rising accents one would expect more words with measurement points HL in the first and more with LH realisations in the second condition. Such a division did not emerge in section 2.4.1. Instead, the LHL combination was most the frequent category within all conditions, so that it seems adequate to describe differences as more or less complete realisations of this one pattern (within the word in question).

As for **Qu3** addressing shape and tonal specification, the present analysis assumes that the underlying pattern is falling, determined by H_pL_p phrase tones (**H3**). This suggestion can be compared to previous analyses in terms of L+H* or LHL accents (Välimaa-Blum, 1988; Suomi et al., 2008, respectively). Figure 2.24 gives annotations based on the three different accounts for an example utterance.[10]

In this context, it is particularly interesting that the values measured as L2 were lowered in narrow focus. In contrast, L1 did not exhibit lowering in narrow focus, but actually had higher values along with the peaks. In the

[10] The targets representing LHL accents are localised according to the phonetic alignment described by Suomi et al. (2008). The authors describe accents as associated with the primary stressed (initial) syllable, but do not designate one of the targets as a starred tone.

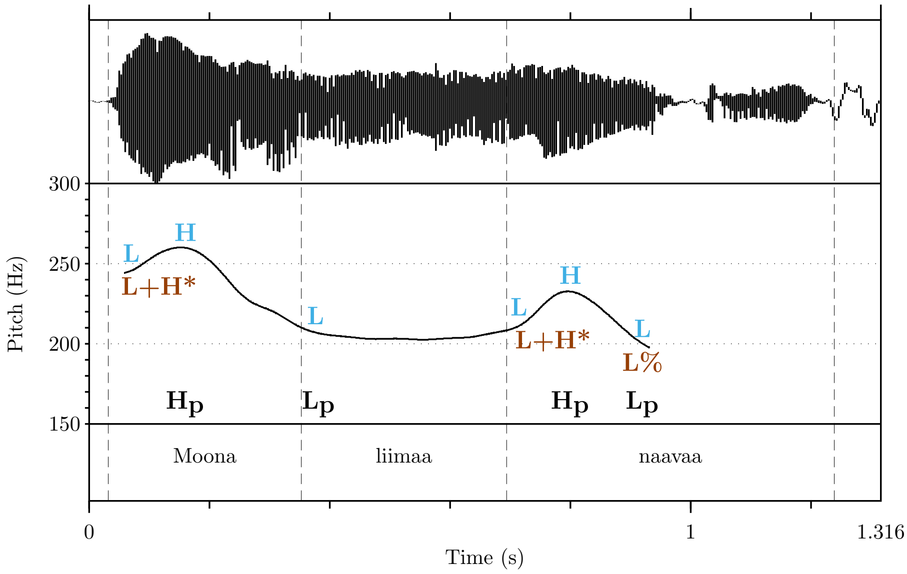

Figure 2.24: Comparison of existing phonological analyses. In light blue (top labels): Annotation based on Suomi et al. (2008). In brown (middle labels): Annotation based on Välimaa-Blum (1988, 1993). In black (bottom labels): Annotation based on analysis suggested here. Sentence *Moona liimaa naavaa* 'Moona glues lichen', broad focus.

present approach, this is expected as H and L2 are assumed to correspond to the phonological targets H_pL_p. It was therefore predicted that these two would be scaled to result in larger pitch ranges for narrow focus constituents (see section 2.3.2, also recall the illustration in Figure 2.7). In contrast, this finding contradicts the assumption of only an L+H* (or rarely L*+H) accent suggested by Välimaa-Blum (1988, 1993). In her account, the fall after the peak is attributed to the L tone of the following accent. Thus, what is measured as L2 here would not be seen as an independent target, but as belonging to the next word's tonal specification (or a low boundary tone in final position). The narrow focus lowering of L2 is difficult to reconcile with such an analysis, since it would need to be interpreted as a lowering of the post-focal word's leading tone. The pitch range of the narrow focus constituent would thus be expanded by manipulating its own H target and the next L target associated with another constituent. It seems much more straightforward to analyse the lowered L2 target as associated with the narrow focus constituent itself, as assumed here. Comparing the two existing accounts of Finnish accents, as L+H* (Välimaa-Blum, 1988, 1993) or as LHL (Suomi et al., 2008), the second one is thus favoured by the data presented here. However, this thesis suggests to further simplify the LHL specification to include only the two tonal targets involved in the pitch range expansion, resulting in a bitonal specification H_pL_p.

Another starting point for questioning that the points measured as L1 were realisations of independent tonal target was the observation that within words, the correlation between the f_0 of L1 and H was better than the correlation between L2 and H (see section 2.4.2). A further reason for the reanalysis as H_pL_p came from the correlation between the two pairs of neighbouring low tones measured in a sentence—i.e. between L2 of the subject and L1 of the verb on the one hand and between L2 of the verb and L1 of the object on the other hand. The values in the two pairs were very similar, indicating a frequent coincidence of what would be described as two neighbouring L targets belonging to different LHL accents in Suomi et al.'s (2008) account. Also, pitch on unfocused given words was not generally flat (see section 2.4.1) as would be expected if they were interpolated from two neighbouring words with LHL tones (independent of whether these tones would be analysed as phrase tones or accents).

Even the time-normalised mean values plotted in Figure 2.6 above strongly indicate that the rise to the peak of the object predominantly already started on the preceding verb and also the point representing L1 of the narrow verb focus condition seems to be part of a rise already starting on the subject rather than a low f_0 turning point. Utterances exemplifying the two cases appeared in Figures 2.10 and 2.25. In Figure 2.10, the points identified as L2 of the verb and L1 of the object do clearly not represent realisations of two different tonal targets, but the rise to the object peak already starts on the verb. In fact, one might argue that the rise to the verb peak starts on the preceding word, and this is clearly the case in Figures 2.25.

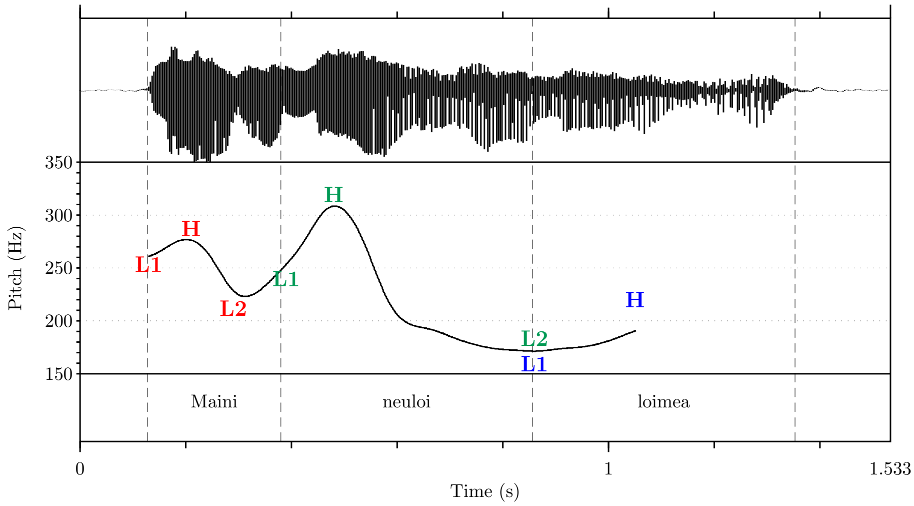

Figure 2.25: Pitch rise to the verb peak already starting on the subject. Sentence *Maini neuloi loimea* 'Maini knitted a blanket', contrastive verb focus.

However, it has to be mentioned that while the points measured as neighbouring L2 and L1 often coincided, they did not always do so. A low plateau can clearly be seen in Figures 2.2 and 2.24. Assuming that the two tonal contours on the subject and the object are not realisations of LHL, but only of H_pL_p targets, the stretch of flat—not interpolated—pitch would require a separate explanation—irrespective of whether the contour is assumed to be shaped by accents, phrase tones or both. The plateau can possibly be viewed as a result of tonal spreading, as illustrated for Figure 2.24, repeated here in Figure 2.26. Conversely, an L+H* analysis would probably assume a leftward spreading of the second accent's leading L tone (a further option is a description as a kind of sagging transition).[11]

2.5.3 Effects of quantity

Most investigated parameters showed some effects of quantity, although the influence was usually less pervasive than that of focus condition or position, as summarised below.

Tonal contour was more often subjectively judged as having an early peak for quantity 2 words (see section 2.4.1).

[11]This analysis mine and is not given by Välimaa-Blum (1988, 1993), who simply states that finite verbs are accentless. In her examples, realisation of pitch on the verbs is usually flat, but may sometimes carry a fall from the preceding word.

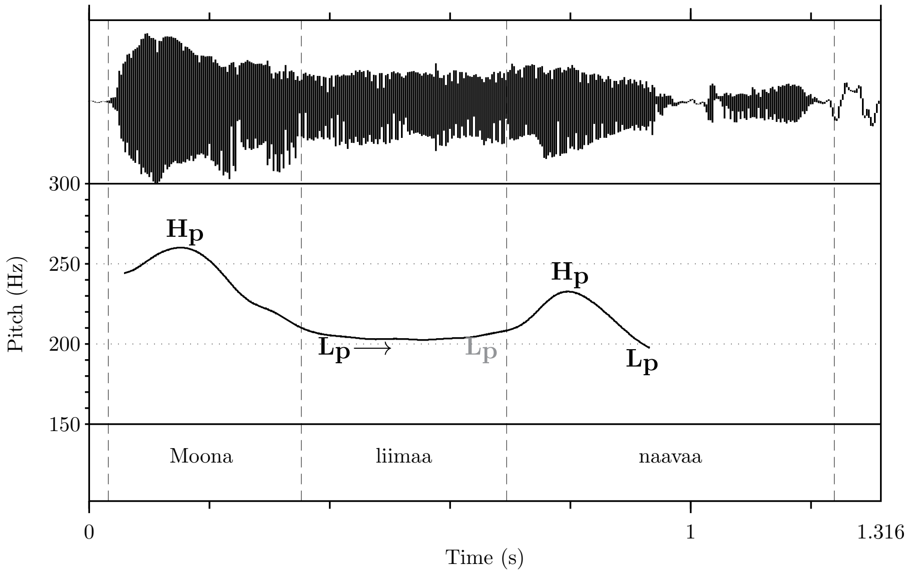

Figure 2.26: Annotation assuming uniform H_pL_p tones with tonal spreading of the L_p. Sentence *Moona liimaa naavaa* 'Moona glues lichen', broad focus (= Figure 2.24).

Pitch scaling of f_0 maxima was higher in quantity 2. Additionally, narrow focus lowering of L2 had a larger effect than in quantity 1 (see section 2.4.2).

Word duration was longer for words containing quantity 2 vowels, especially in position 2 (see section 2.4.3).

Alignment of all three f_0 measurement points differed between the quantities, although only in some positions and / or focus conditions (see section 2.4.4).

Voice quality was more often non-modal on the first syllable of narrow focus words when it contained a quantity 2 vowel. Also second syllables of narrow information focus words were more often non-modal in quantity 2 (see section 2.4.5).

Pauses did not show an effect of quantity (see section 2.4.6).

This section generally explains the effects as due to the durational variation induced by the quantity manipulations, i.e. long vowels affording more space than short ones. Quantity had an overall effect on **word duration**, which was longer in quantity 2, as predicted in section 2.3.3. Additionally, an interaction pointed to a larger quantity effect in position 2. This was likewise expectable, since the second position was the only place where first and second syllable vowel quantity were varied simultaneously. Contrary to the predictions in section 2.3.2 and the previous literature, quantity effects were also found for **pitch scaling**. Interestingly, this was only the case for H and L2 points, which generally corresponded to the phonological targets H_pL_p under the present analysis. The findings thus fit the interpretation that the durational realisation of long or short quantities affected pitch scaling to a limited extent by providing more or less space for reaching tonal targets. Likewise, this section assumes that duration can explain the observed quantity differences for the segmental **alignment**. For all three f_0 measurement points, if a quantity difference appeared, more f_0 points were detected on the first syllable vowel in quantity 2 than in quantity 1. In other words, when the duration of the first vowel was longer it was more likely that a local minimum or maximum was identified during its duration. Similarly, the interaction of quantity and narrow focus condition for the **voice quality** of first syllables means that when the first syllable was long, non-modal phonation of the second syllable already began on the preceding first syllable. Put differently, the stretch of modal voice quality before the beginning of the focus-marking creak was sometimes shorter than the first syllable. This was more likely when the first syllable contained a quantity 1 vowel and had a short duration. In contrast, more frequent non-modal voice realisations of second syllables in narrow focus quantity 2 words are less easy to explain, but notice that the effect was only significant for information focus.

Seeing quantity differences as due to durations, a question is of course whether in the subjective categorisation of **tonal contours**, peaks were perceived as earlier in words with quantity 2 vowels simply because these words are longer. It might be that an identical peak timing with respect to the beginning of the word—as found by Wiik (1988); Suomi (2007a)—appears to be earlier relative to a longer word. Notice, however, that peak **alignment** in absolute terms as measured from the beginning of the word was considerably later for quantity 2 words in the non-nuclear conditions and only very slightly earlier than for quantity 1 in the nuclear ones (see section 2.4.4, especially Figure 2.16). Regarding question **Qu1**, the crucial point is whether the quantity differences in peak alignment can be explained as non-categorical variation in the realisation of a single tonal pattern or whether the induced variance signifies the presence of categorically different accents. This chapter argues that for an analysis in terms of only one set of—somewhat variable—phrase tones H_pL_p (see **H1**, **H2** and **H3**), but an alternative possibility is sketched in section 2.5.4 below and tested in the next chapter.

Finally, it is possible that some observed quantity effects on the segmental **alignment** of the measured points L1, H and L2 are not purely due to different durations of the first vowel, but support a specification of tonal association in moraic terms, as suggested by Suomi and his colleagues (e.g. Suomi et al., 2003; Suomi, 2007a; Suomi et al., 2008). Ultimately, this question is left somewhat open here.

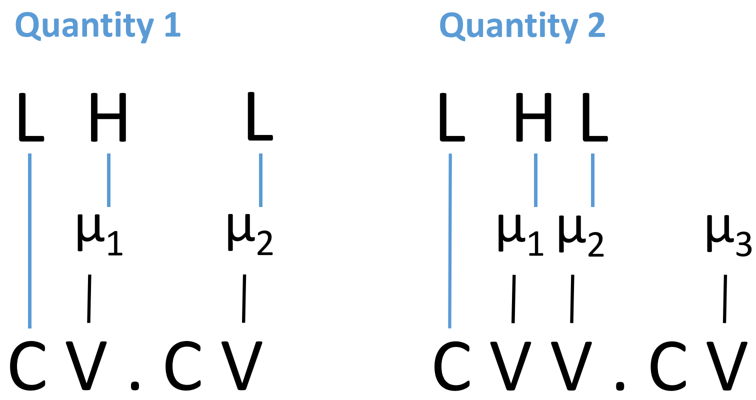

Figure 2.27: Tonal alignment with respect to moras according to Suomi (2007a).

As illustrated in Figure 2.27, Suomi assumes that across quantities, pitch contours can be described as showing stable alignment of three tones: a first L tone aligned with the word onset consonant, a H tone aligned with the end of the word's first mora and a second L tone aligned with the end of the word's second mora (or later according to Suomi (2007b)). In the present data, the measurement point L1 was mostly marked at the expected location. However, recall that L1 was defined as the minimum preceding a word's maximum within the same word, so that rises starting before the beginning of the respective word were not detected here.

For the f_0 maximum, a difference between the quantities occurred in sentence-initial position: More H points were identified on the first syllable vowel nucleus V1 in quantity 2 than in quantity 1, whereas they appeared on the following consonant C2 more often for quantity 1 than for quantity 2 (see section 2.4.4). This can be taken as support for Suomi's (2007a) claim that H is aligned with the end of the first mora. Of course, the precise location corresponding to the end of the first mora was not measured, so that the claim can only be assessed in a somewhat indirect way here. However, one mora is associated with quantity 1 vowels, whereas quantity 2 vowels bear two moras. Thus, while end of the first mora is not registered in itself for quantity 2, one can expect that peaks would be realised well during the first syllable nucleus on the basis of Suomi's account (presumably around the midpoint). For quantity 1, in contrast, an exact alignment to the moraic boundary would fall between the categories V1 (first syllable vowel nucleus) and C2 (onset consonant of the second syllable). Peaks should thus have been measurable either on V1 or on C2 for quantity 1. A higher percentage of C2-alignments would therefore be predicted for quantity 1 than for quantity 2, and this is what was found across all focus conditions for the sentence-initial words.

The f_0 measurement point L2 was mostly found on the second syllable nucleus. Since L2 was defined as the second minimum of the respective word and it was thus not checked whether the pitch fall continued on the following word, this finding is compatible with Suomi's (2007b) finding that L2 is realised around the middle of the third syllable in longer words. However, a part of the data also agrees with Suomi's (2007a) finding of L2 tones being realised towards the end of the second mora. Namely, the second low turning point was measured on V1 more often in quantity 2 than in quantity 1, and thus more often when V1 comprised both the first and the second mora. Quantity 2 words also exhibited more L2 realisations on the intervocalic consonant C2, i.e. 'on the other side of the end of the second mora', in most focus conditions and positions.

2.5.4 An alternative account: the nuclear accent hypothesis

This section sketches an alternative account of the data that will be tested—and refuted—in Chapter 3. This account, referred to as the nuclear accent hypothesis, differs from the assumptions advanced in this chapter with respect to all three research questions, **Qu1**, **Qu2** and **Qu3**.

With respect to **Qu1**, the alternative hypothesis **A1** is that more than one category of tonal realisation exists. The current data do not provide evidence for a complex accent inventory, but at best for a binary distinction. The most likely candidate for a categorical division is termed nuclear vs. non-nuclear accent here and preliminarily designated as H*L vs. LH* (i.e. making reference to accents in line with **A2** and using a tonal configuration differing from the suggestion in **H3**). The suggested difference would be a rather fine-grained one. As the notation indicates, the H tone of both accents would associated with the stressed—i.e. initial—syllable. The hypothesis rests on interpreting the different peak alignments appearing in Figure 2.16, repeated as Figure 2.28, as evidence for categorically distinct accents.

Figure 2.28 suggests that peaks in nuclear conditions (in green) were consistently aligned to the beginning of the first syllable nucleus, while the peak alignment in non-nuclear conditions (in blue) can be described in a more straightforward way with reference to the end of the first syllable nucleus.

In particular, this switch of reference might help to unify differences between the quantities. For nuclear accents, Figure 2.28 indicates that peak timing for both quantities was similar relative to the beginning of the word, in line with the findings of Suomi et al. (2003); Suomi (2005, 2007a,b, 2009); Ylitalo (2009). If any differences appeared, the timing was slightly later for

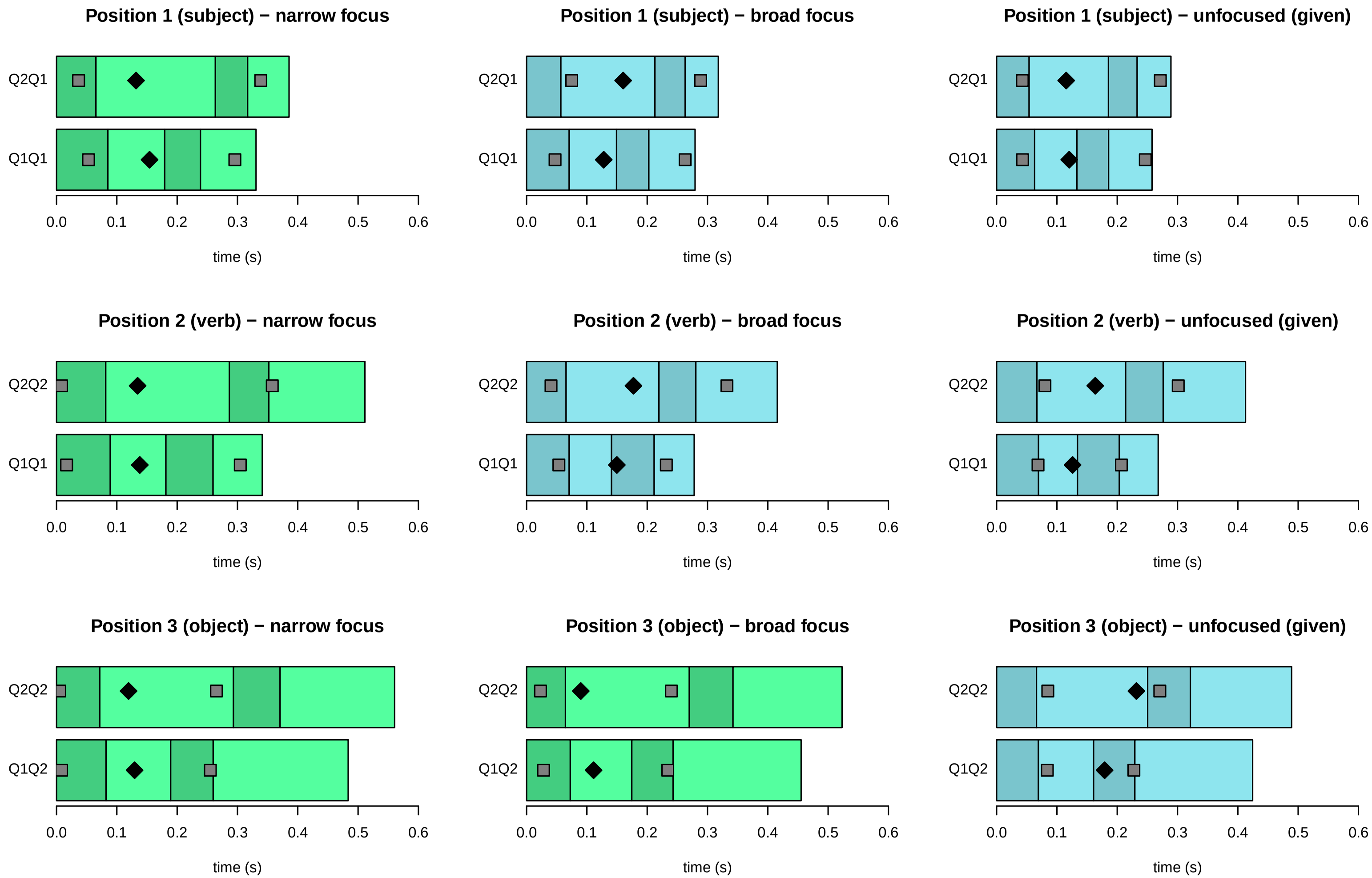

Figure 2.28: Average segment durations and times of f_0 turning points for all experimental conditions (= Figure 2.16).

words with first quantity, as for example for narrow focus subjects. For non-nuclear accents, in contrast, peaks of quantity 2 words appeared clearly later as seen from the beginning of the word. Thus, in comparison to the nuclear accent realisations, the peak seemed to have moved backwards a lot more in quantity 2 than in quantity 1. However, these clear differences only emerge when alignment is considered as relative to the beginning of the word or first syllable nucleus.

A more unified account of the timing for non-nuclear accents can be given assuming that the segmental landmark targeted by H alignment in these cases was the end of the first syllable nucleus. Relative to the end, the distance was larger in quantity 2, which implies that variation in quantity 1 was more constrained due to shorter vowel duration. In quantity 1, peaks in non-nuclear accent condition appeared either shortly before or shortly after the boundary between the first syllable vowel and the following consonant. On the one hand, the H tone needs to be realised during the first syllable or in close enough vicinity to mark its association with the stressed syllable. On the other hand, a too large distance from the end of the syllable would risk inducing a perception of the peak as aligned to the syllable or nucleus beginning instead of its end and thus of a timing corresponding to the nuclear accent. In quantity 2, there is more freedom in this respect. Since the vowel is long, the peak can have a considerably larger distance to the syllable's end than in quantity 1, while still being clearly realised during the second half of the syllable nucleus and thus aligned to its end rather than to its beginning.

In sum, the pattern emerging from Figure 2.28 would be that H tones were realised (relatively) close to the beginning of the first syllable vowel for nuclear accents (tentatively formalised as an H*L accent), while non-nuclear accents showed H tones (relatively) close to the end of the first syllable vowel (tentatively formalised as LH*). The reason for referring to these two different alignment patterns as 'nuclear' vs. 'non-nuclear' accents is that the sentence-final objects showed the same peak timing in broad focus sentences and when they were in narrow focus themselves. This suggested borrowing the term from descriptions of Germanic languages, where broad focus is also regularly marked by what is referred to as nuclear accent on the direct object (the focus exponent). More generally, this phenomenon has been referred to as focus projection (e.g. Chomsky, 1971; Gussenhoven, 1983; Krifka, 1984).

The nucleus is the essential component of an intonation contour in British school descriptions, realised on the last accented syllable of the phrase (e.g. O'Connor and Arnold, 1973; Cruttenden, 1997). Although Pierrehumbert (1980) did not assign nuclear accents a theoretically distinguished status in her description of American English, she continued to use the term and later work within the AM framework has argued for substantial difference between nuclear and other pitch accents (e.g. Gussenhoven, 2004). Nuclear accents can usually be identified as the final and most prominent accent in a phrase and their final position is sometimes seen as their defining characteristic (e.g. Beckman and Pierrehumbert, 1986, p. 295). An understanding following Gussenhoven (1983) will be assumed here: Based on the information structure of the sentence, specifically the division into focused and non-focused (background), accents are assigned in a regular way and the last accent assigned to a / the focused constituent of the phrase is the nuclear accent (Gussenhoven, 1983, especially p. 382).

Crucially, adopting the term nuclear accent here rests on the behaviour of the objects, for which the same average tonal alignment pattern was observed in broad and narrow focus. This similarity could be accounted for using the concept of nuclear accent as outlined above. It could thus be said that in Finnish, the rightmost word of a focus constituent (the focus exponent), receives the nuclear accent of the sentence, while the other (pre- or post-focal words), can receive non-nuclear accents. In the contexts eliciting narrow focus, only one word was focused in the answer and this word thus carried nuclear accent. In the case of broad focus, however, the focus constituent containing the answer to the question consisted of the whole sentence. The object, as the rightmost word of this focus constituent, was its focus exponent and thus received the nuclear accent. This is illustrated in (18), where boldface indicates the location of the nuclear accent.

(18) a. Narrow focus on the subject: $[\mathbf{S}]_{\mathrm{F}}$VO

b. Narrow focus on the verb: S$[\mathbf{V}]_{\mathrm{F}}$O

c. Narrow focus on the object: SV$[\mathbf{O}]_{\mathrm{F}}$

d. Broad focus: $[\mathrm{SV}\mathbf{O}]_{\mathrm{F}}$

Assuming that (18d) captures the findings of the current chapter, they could then be summarised by saying that nuclear accents are marked by an earlier peak (tentatively notated as H*L vs. non-nuclear LH*; section 2.4.4), long duration (section 2.4.3), a large pitch fall (section 2.4.2), as well as optionally by a pause (section 2.4.6) and / or non-modal phonation (section 2.4.5) following them. This description nicely accounts for the similarity of the broad focus objects and the narrow focus words in the present data set and additionally provides a common denominator for the other means of focus marking. In fact, the suggestion of the nuclear accent as a distinct category capitalises on the observation that differences between broad and narrow focus realisations of sentence-final objects were minimal for tonal alignment and voice quality. It concludes from this that the prosodic realisation of the sentence-final word is determined by the same—nuclear—accent in both cases. However, the account does not cover all the observed effects. Also in final position, a pitch range expansion did occur in narrow focus compared to broad focus, which is not predicted by the nuclear accent hypothesis. Moreover, the special behaviour of the sentence-medial verbs would lead to the separate assumption that they were often, but not always unaccented outside narrow focus.

A further possible problem for the nuclear accent hypothesis is the pitch valley before the assumed H*L accents. To take only one example, consider the pitch dip between the peaks of subject and verb in Figure 2.14. According to the suggested analysis, the narrow focus verb would carry the nuclear accent H*L. If the preceding non-nuclear accent on the subject were correctly annotated as LH* it would be very difficult to explain the pitch fall following it. The same problem occurs for practically all example utterances depicted in the current chapter, not to mention the example in Figure 2.5, in which the last pitch peak of the sentence—located on the verb—clearly does not match the description of the nuclear accent given above. A possible solution would be to appeal to phrasing. In general, the assumption of (at least) two categorically different accents and an account of phrasing like the one presented in (17) are not mutually exclusive. Most likely, it would then make sense to say that p-phrases are marked only by a final boundary tone L, while the contrast between nuclear and non-nuclear accents would be formulated as one between H* and LH*. Presently, this is little more than speculation. Therefore, the following chapter will scrutinise the measured peak alignment again to see whether it warrants the proclamation of two categorically different patterns. To this end, it will statistically test concrete predictions generated from both the nuclear accent hypothesis and the uniformity hypothesis.

2.6 Summary

This chapter discussed the results of Production Experiment I, investigating the effects of vowel quantity, focus condition and position / grammatical function on the tonal contours of all words in short SVO sentences. It considered the identifiability of maximally three points—L1, H and L2—their timing and pitch scaling, as well as the additional parameters duration, voice quality and pause occurrence. For most investigated parameters, effects of all three factors were observed.

In analysing the results of this experiment, the chapter referred to the prosodic hierarchy as sketched in section 1.4. It characterised the p-phrase as marked by H_pL_p tones and tentatively suggested that final lengthening occurred to some degree for p-phrases. It also referred to i-phrases right-edge aligned with focus constituents and marked by a L_p tone, final lengthening, pauses and final non-modal voice quality.

The analysis submitted that all observed realisations of the SVO sentences can be modelled using a uniform specification of p-phrase tonal contours as falling, i.e. defined by H_pL_p tones. Importantly, very little evidence for the relevance of contrasting accents emerged.

Chapter 3

Refuting the nuclear accent hypothesis

3.1 Introduction

The previous chapter proposed an analysis of Finnish tonal contours as shaped by phrase tones. However, it also mentioned a possible alternative account involving two different types of pitch accents. Thus, it concretely outlined two possible hypotheses on the basis of a production experiment. First, that Finnish tonal contours are all of the same category, i.e. uniform as assumed in previous literature—although it modelled these contours not as realisations of a uniform accent, but of phrase tones. Second, it introduced the competing possibility that will be referred to here as the nuclear accent hypothesis. It posits that there is a categorical difference between nuclear accents, i.e. the last accents in focus constituents (focus exponents), and all other accents. It assumes that the peaks of nuclear accents are consistently aligned to the beginning of the first syllable nucleus and those of non-nuclear accents to its end, which is preliminarily designated as H*L vs. LH* accents.

The current chapter statistically tests the nuclear accent hypothesis against the hypothesis of uniform alignment. To this end, it examines whether the postulated different peak alignments are reliably found in a carefully edited version of the data set. Concrete predictions for the effects of the experimental conditions on different measures of alignment are derived from both the nuclear accent hypothesis and the uniformity hypothesis and subsequently tested.

Two aspects of this deserve further elaboration here, the editing of the data set and the choice of measures of alignment. The preceding chapter illustrated a need for data editing by showing that the f_0 maxima of some words in the recorded materials are not actually peaks (see section 2.4.2). While these should thus be excluded, excessive data cleaning is risky, especially if it is based on subjective decision. The issue was therefore addressed here by using an automatic data editing method, which is introduced in section 3.2.1.

As for the measures of alignment, several suggestions are present in the literature. The choice seems to depend on underlying theoretical assumptions to some extent, but also on the investigated language. Schepman et al. (2006) compare different options in detail for materials that are in some respects similar to the ones used here, namely for Dutch rising-falling nuclear accents on words with either phonologically short or long vowels in the initial stressed syllable (the article includes a comparison with similar data on pre-nuclear accents originally analysed by Ladd et al., 2000). Their methodological conclusions are taken as guidance here (see the discussion in section 3.2.2).

Following this analysis of peak alignment in section 3.2, sections 3.3 and 3.4 present a pilot perception study and an analysis of intensity, respectively, which likewise have a bearing on the nuclear accent analysis.

3.2 Peak alignment in the data from Production Experiment I

3.2.1 Data and data editing

The data used were the same 947 short SVO sentences as in the previous chapter, in which the factors (first syllable vowel) quantity and information structure were independently varied, while grammatical function was correlated with position in the sentence due to invariable word order (for a documentation of the elicitation method see Chapter 2, section 2.2.1). For these materials, the f_0 and time of each word's pitch maximum, as well as of two minima left and right of it had been measured among other parameters. Before analysing the peak alignment, measurements from pitch maxima that were not 'real peaks' were sorted out. This process took into account the preceding chapter's observation that pitch movements frequently exceeded a single word, so that for example the rise to the object peak started on the verb already. Two kinds of data points were eliminated.

First, pitch maxima that were not f_0 turning points, but part of a rise to a peak on the following word or of a fall from a peak on the preceding word should clearly not be considered in an evaluation of peak alignment. The occurrence of these kinds of contours has been illustrated in the previous chapter, where their frequency was gauged by subjective judgement. Here, objective criteria were adopted to identify 'fall from' and 'rise to' contours. As in all the automatic evaluations characterised in this section, criteria for 'early' and 'late' were to be located in the first and last third of the word, respectively, while a perceptually relevant pitch movement was taken to be at least 0.8 st (on the choice of this criterion see Appendix C, also see Appendix F for results of applying different criteria).

Words with the ‘fall from’ type of contour were assumed to have...

1. an early f_0 maximum,
2. a maximum lower than the preceding word’s maximum
3. no perceptually relevant pitch rise preceding the maximum within the word.

As an example, consider Figure 3.1, where the originally measured f_0 points for each word are labelled as L1, H and L2, respectively, in red for the subject, green for the verb and in blue for the object. These original measurements detected a pitch peak close to the end of the verb, followed by a minimal fall, and identified the object’s f_0 maximum as being at its beginning. Still, whatever analysis of Finnish intonation one might choose, it seems obvious that these points do not represent three phonological targets. The utterance shows two rising-falling pitch movements, one on the subject (with a minimal rise only) and one that spans the VP. Therefore, only the higher H point was selected to represent the latter movement in the current data set, H of the verb, which was the local f_0 maximum and turning point. The lower H of the object, in contrast, was determined to be part of a ‘fall from’ contour and not a peak itself according to the three criteria given above.

Mirroring the criteria for identifying ‘fall from’ contours, words with the ‘rise to’ contour type were assumed to have...

1. a late f_0 maximum,
2. a maximum lower than the following word’s maximum
3. no perceptually relevant pitch fall following the maximum within the word.

An example is given in Figure 3.2. No L2 value was detected on the subject of this sentence, for which the maximum pitch is reached at the end of the word. This maximum, indicated by the red letter H, is located during the rise to the verb’s peak and was therefore sorted out. The pitch maxima of the verb and the object of the same sentence were included in the data set.

Second, maxima which differed from both neighbouring minima by less than 0.8 st were considered to represent perceptually insignificant pitch movements and were therefore sorted out. For example, the small movement visible around the middle of the subject in Figure 3.2 would not have passed this criterion, but the pitch movements on the object in the same sentence spanned more than 0.8 st. However, the previous examples have illustrated that it is not enough to check one word’s pitch range to see whether its peak was part of a significant movement, as rises frequently started on a preceding

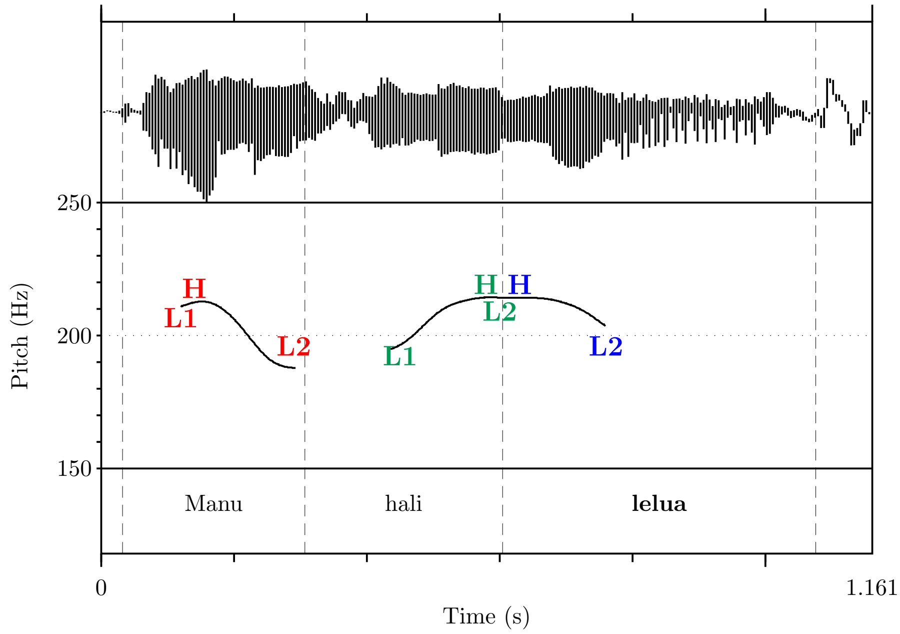

Figure 3.1: Example of a 'fall from' contour on the sentence-final object (in boldface). Sentence *Manu hali lelua* 'Manu hugged a toy', broad focus.

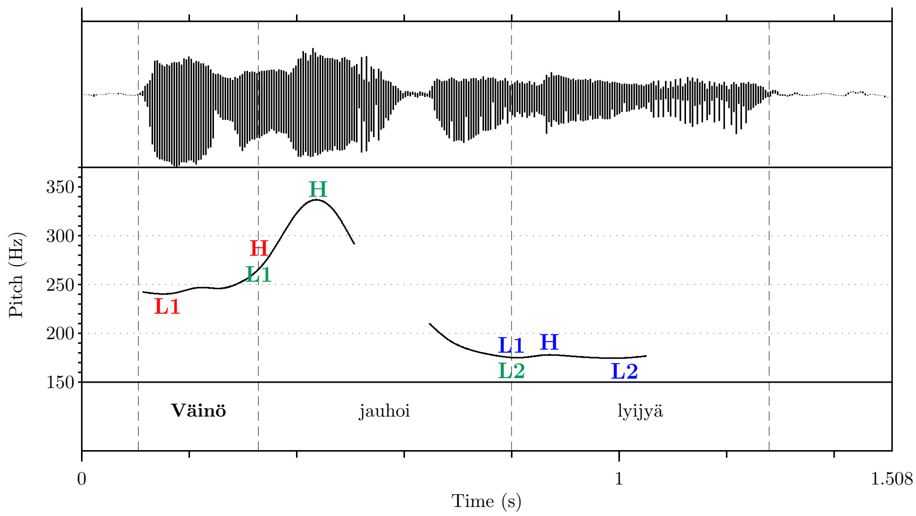

Figure 3.2: Example of a 'rise to' contour on the sentence-initial subject (in boldface). Sentence *Väinö jauhoi lyijyä* 'Väinö ground lead', narrow corrective focus on the verb.

word and falls often continued on the next one (as was repeatedly observed in Chapter 2). For example, the peak on the verb in Figure 3.1 is followed by a fall that is mostly localised on the object. Thus, before evaluating whether the peak was part of a pitch movement with a sufficiently large range, an automatic process determined the two minima to be used in this comparison. It identified movements that started or ended on the preceding or following word by finding 'fall from' and 'rise to' contours as outlined above, as well as cases like illustrated in Figure 3.3. In this utterance, the beginning of the rise to the object's pitch maximum was determined to be the point measured as the verb's L2 (in green) instead of the minimal perturbation originally identified as the object's L1 (in blue). Note that the size of the rise was the decisive factor here, as the sentence ended in non-modal voice, so that no final pitch fall could be measured.

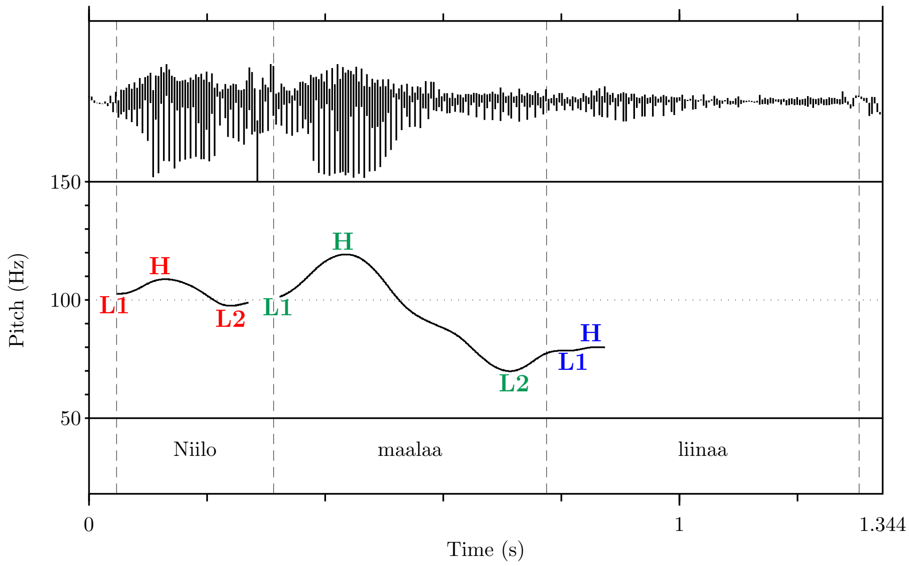

Figure 3.3: Rise to the object's peak starting on the verb. Sentence *Niilo maalaa liinaa* 'Niilo paints a cloth', narrow corrective focus on the verb.

The R (R Development Core Team, 2010) code used for all of these automatic classifications is documented in Appendix E. It sorted out 39 subjects as 'rise to' contours and 20 as having flat pitch with the f_0 maximum differing by less than 0.8 st from the neighbouring minima, while for nine subjects, non-modal voice prevented pitch measurements during more than an extremely short stretch. Of the verbs, the code identified 117 as 'rise to' and 143 as 'fall from' contours, as well 31 as carrying only small pitch movements; sufficient measurement was not possible in 90 cases. Of the objects, 86 were classified as 'fall from' contours, 44 as having no relevant pitch movements and 225 had to be excluded due to an insufficient amount of measurable values. This can be compared to the author's judgements in Chapter 2 (see Table 2.4), but apart from the classification of 35 subjects as 'rise to following' there, the correspondence is not close. One reason for this is a factual overlap—e.g., a 'rise to' movement might be smaller than 0.8 st—due to which the order in which the code in Appendix E checked for a match with each of the categories influenced their relative frequency. This overlap is not relevant to the outcome, as f_0 maxima matching any of the categories were eliminated. Note, however, that the current method is more conservative than data cleaning on the basis of the subjective judgement represented in Table 2.4 would have been, where an overall greater number of words was judged not to have an own peak. Also, the present method is based on objective criteria, lowering the risk of experimenter bias.

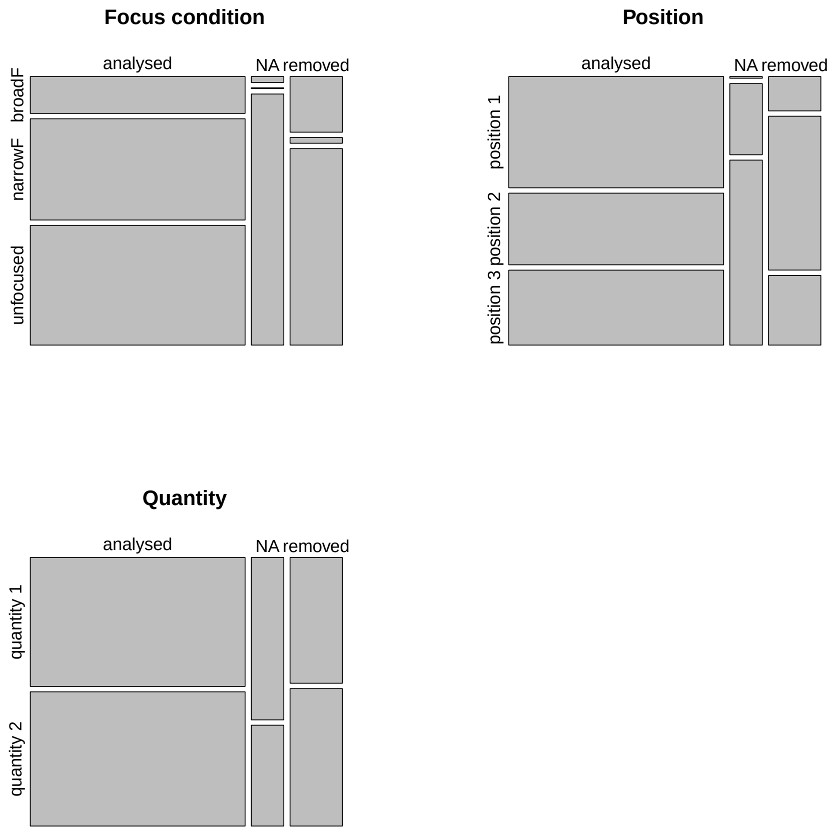

Figure 3.4: Data editing and distribution of experimental conditions: focus condition (top left), position (top right) and quantity (bottom left).

Table 3.1: Distribution of data points (absolute numbers) for the analysis of alignment across different combinations of conditions. CF = contrastive focus, IF = information focus.

		Position 1	Position 2	Position 3
Quantity 1	Broad focus	62	46	45
	Narrow CF	68	67	67
	Narrow IF	64	66	64
	Unfocused (CF)	122	58	41
	Unfocused (IF)	123	60	45
Quantity 2	Broad focus	54	38	46
	Narrow CF	66	68	67
	Narrow IF	68	68	67
	Unfocused (CF)	128	49	68
	Unfocused (IF)	124	46	82

Figure 3.4 compares the balance of conditions in the resulting data set (bar 'analysed') and for the data removed by the editing process (bar 'removed'; the bar 'NA' represents words where f_0 measurements were impossible due to non-modal voice quality). Table 3.1 shows the exact number of analysed measurements in each condition. Overall, the edited data set contained 291 words in broad focus, 800 in narrow focus and 946 in unfocused condition, while editing removed 107, 11 and 378, respectively. Recall that originally, broad focus realisations were as frequent as narrow information focus and contrastive focus realisations, respectively, while the unfocused conditions were double as frequent, since in every sentence with a narrow focus constituent the two other words were unfocused (and contextually given). This imbalance was somewhat rectified by the data editing procedure, which removed only 1% of measurements in narrow focus, but 26% in broad focus and 23% in unfocused condition (in addition, 18% of unfocused words had no measurable peak to begin with, compared to 2% in broad focus and 0% in narrow focus). Regarding the positions, Figure 3.4 illustrates that the procedure sorted out most data points for the sentence-medial position (31% of the recorded verbs vs. 7% and 14% of subjects and objects, respectively), while 23% of sentence-final objects had no measurable peak due to voice quality (vs. 0% of subjects and 9% of verbs). As a result, the analysed data set contained 879 words in position 1, 566 in position 2 and 592 in position 3. Finally, Figure 3.4 suggests that both quantities were equally affected by editing, but f_0 was not measurable more often for the shorter words in quantity 1 (13%, vs. 8% for quantity 2). As a result, the edited set contained 998 quantity 1 words and 1039 quantity 2 words.

Altogether, 2037 peaks were retained for analysis in the edited data set. Their alignment was evaluated with linear mixed-effects models, which are known for their ability to handle unbalanced data sets well (R Development Core Team, 2010; Baayen, 2008). As in the previous chapter, effects with t $> |2|$ were interpreted as significant (see section 2.2.4).

3.2.2 Hypotheses

This section details the two hypotheses on peak alignment, the nuclear accent hypothesis and the uniformity hypothesis. It then determines three appropriate measures of alignment to compare the hypotheses and derives concrete predictions for the outcome of linear mixed-effects models of the measures. The two hypotheses are listed below.

The nuclear accent hypothesis Peaks of nuclear accents are aligned to the beginning of the first syllable nucleus, peaks of non-nuclear accents are aligned to the end of the first syllable nucleus.

The uniformity hypothesis The timing of peaks is uniform across quantities and focus conditions. In particular, there is no significant difference between words in nuclear and non-nuclear position. All peaks are realised at a fixed distance from the beginning of nucleus.

A schematic illustration of the two hypotheses to be compared appears in Figure 3.5, with the nuclear accent hypothesis in the left panels and the uniformity hypothesis in the right panels. It is overall similar to Figure 2.16 illustrating mean segment duration and tonal timing in the previous chapter and can be interpreted similarly. However, it collapses several experimental conditions and distinguishes only two cases: Contexts where nuclear accents appear according to the nuclear accent hypothesis (words in narrow focus and sentence-final objects in broad focus) and those where non-nuclear accents appear (all other combinations of the factors position and focus condition). The segment duration bars—again distinguishing quantity 1 (lower bars) and quantity 2 (upper bars)—in each of the two simplified conditions depict actually occurring means, namely for the verbs in narrow and broad focus, respectively.[1] In contrast, the timing of the peaks as indicated by the black diamonds illustrates the predicted means for the edited data set evaluated here. While the two hypotheses are different interpretations of the findings in the previous chapter, the illustration in Figure 3.5 is an idealisation of these findings based on these respective interpretations. The results for the edited data set used in the present chapter should approach either of the illustrated

[1] Recall that vowel quantity varied in both syllables for verbs, while only first syllable quantity changed for subjects and objects (see section 2.2.1). Verb durations were therefore chosen for visual clarity and simplicity of labelling, but for the evaluation of the hypothesis, second syllable quantity is irrelevant.

idealised timing patterns more closely than Figure 2.16, since measurements of maxima that are not peaks were excluded.

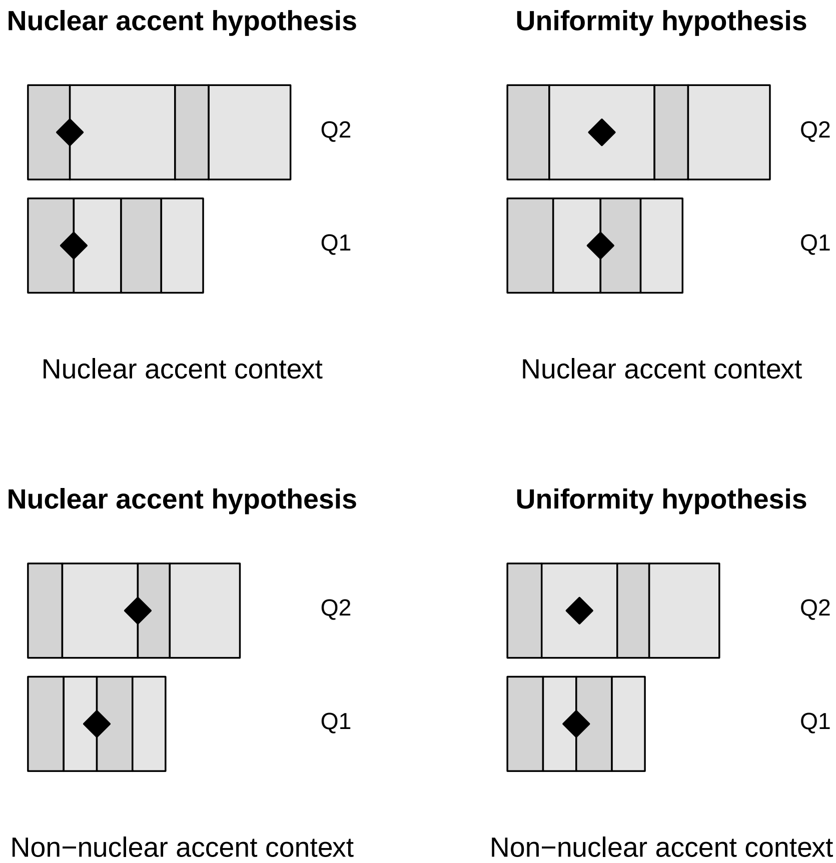

Figure 3.5: Schematic illustration of expected peak alignment in different contexts according to the nuclear accent hypothesis (left) and according to the uniformity hypothesis (right).

In selecting measures for statistically assessing the hypotheses, the discussion in Schepman et al. (2006) was taken as instructive. After comparing results for different measures of alignment for their data, the authors draw two methodological conclusions. First, they argue that tonal alignment should rather be measured relative to a segmental landmark than relative to another tonal event, as the former resulted in more consistent findings and a better model fit for their data. Second, Schepman and her colleagues find that "it is better to express alignment relative to a nearby acoustic landmark than a more distant one" (Schepman et al., 2006, p. 22), since more distant landmarks lead to greater variance, which is likely to introduce uninformative correlations.[2]

Following these recommendations, the current analysis thus concentrated on measures of absolute distances between peaks and relevant segmental landmarks. The previous chapter found that most peaks occurred during the first vowel, i.e. the first syllable nucleus (see section 2.4.4). Thus, the beginning and end of this vowel are the nearest segmental landmarks. To test the nuclear accent hypothesis, peak timing was analysed relative to both segment boundaries, since the hypothesis claims that both of them are targets of alignment. The uniformity hypothesis as stated above refers exclusively to the beginning of the nucleus. Of course, one could hypothesise uniformity relative to other landmarks, as well. Here, the beginning of the nucleus was chosen as a nearby reference point which corresponds reasonably well both to the claims of uniformity in previous literature (Wiik, 1988; Suomi, 2005) and to the findings presented below. The next chapter discusses other targets for alignment and on the basis of more suitable data confirms the decision for the beginning of the first syllable nucleus as the most consistent point of reference (see especially section 4.4.4.1). Since the focus of the current chapter is to compare the nuclear accent hypothesis and the uniformity hypothesis, section 3.2.3 analyses alignment in terms of the two segmental landmarks most relevant to this comparison. The two measures were calculated in the same way as done by Schepman et al. (2006) (who analysed the Dutch nuclear accents in their data in terms of the distance from the end of the nucleus), that is by subtracting the time of the landmark from the time of the peak. The two measures of alignment used here were thus:

Distance from beginning of nucleus Distance of the peak from the beginning of the first syllable nucleus in milliseconds (ms). Calculated as: *time of peak (H) - time of beginning of nucleus.* The measure is positive when the peak is realised during the nucleus of the first syllable or later. The larger the distance, the bigger the value is.

Distance from end of nucleus Distance of the peak from the end of the first syllable nucleus in ms. Calculated as: *time of peak (H) - time of end of nucleus.* The measure is negative when the peak is realised during the nucleus of the first syllable or earlier (during the word onset). The earlier the peak, the more negative the value is.

[2]The third methodological question that Schepman et al. (2006) address, but ultimately leave open, is whether expressing alignment in proportional measures (e.g. 'peaks were on average aligned to the middle of the syllable') is to be preferred over using absolute ones (e.g. 'peaks were on average aligned 30 ms after the beginning of the syllable'). However, the results presented in section 3.2.3 indicate that the alignment of Finnish peaks cannot be described as a constant proportion of a prosodic unit like word, syllable or syllable nucleus. An analysis in terms of a proportional measure was therefore not tested here.

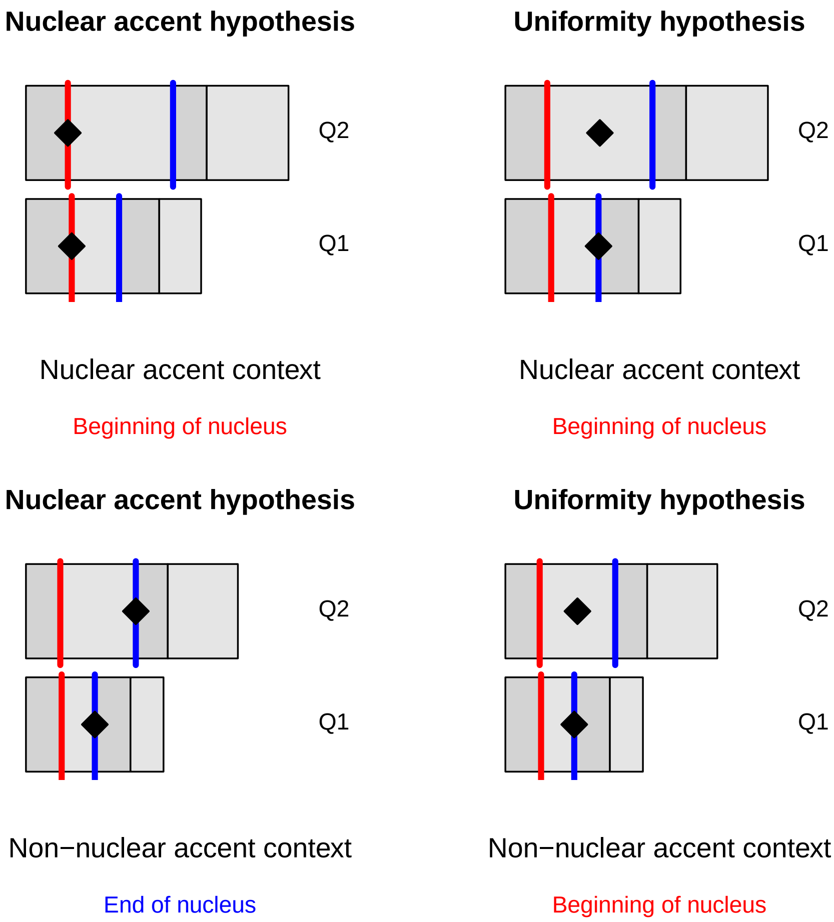

Figure 3.6: Schematic illustration of expected peak alignment relative to two segmental landmarks: 1. beginning of the first syllable nucleus (red lines), 2. end of the first syllable nucleus (blue lines). Left panels: according to the nuclear accent hypothesis. Right panels: according to the uniformity hypothesis (= Figure 3.5 with added visualisation of alignment targets).

The two reference points are marked by red and blue lines, respectively, in Figure 3.6. This figure serves to outline the predicted outcomes for the analyses of the two measures. The nuclear accent hypothesis states that there is a categorically different alignment for nuclear accents (top left panel) and for non-nuclear accents (bottom left panel): Peaks of nuclear accents are aligned to the beginning of the first syllable nucleus, while peaks of all other accents appear at the end of the first syllable nucleus. An important aspect about this shift in the point of reference is how it relates to the quantities. Peak alignment for both quantities is very similar in the nuclear accent condition (in fact, exactly the same in this idealised representation). For the non-nuclear accents, alignment seems to be quite different in quantity 1 than in quantity 2 at first glance. However, according to the nuclear accent hypothesis it is also the same when seen from the end of the first syllable nucleus. However, due to the difference in vowel duration, the distance from the beginning of the nucleus (and of the word) is clearly larger in quantity 2. Conversely, the distance between the peak and the end of the nucleus is larger for nuclear accents than for non-nuclear accents according to this hypothesis. Again, the absolute difference between nuclear and non-nuclear alignment is larger for quantity 2, which has a longer syllable nucleus duration.

Thus, according to the nuclear accent hypothesis the following two predictions should hold of the linear mixed-effects models for both measures of alignment:

N-P1 Nuclear accents have earlier peaks than non-nuclear ones.
⇒ There is a main effect of nuclear condition.

N-P2 The distinction between nuclear and non-nuclear accent has a larger absolute effect for quantity 2 than for quantity 1.
⇒ There is an interaction of quantity and nuclear condition.

In contrast, the uniformity hypothesis claims that alignment as seen from the beginning of the nucleus (red lines) is the same in words that are final in a focus constituent as in those that are not. In line with this, the top and the bottom right panel in Figure 3.6 only differ in their segmental durations, which are larger in 'nuclear accent' contexts (for the findings on duration, see section 2.4.3). Likewise, the two quantities are supposed to have the same tonal contour realised on different segmental durations. Since the hypothesis assumes that alignment is uniform relative to the beginning of the first syllable nucleus, differences in the distance from the end of the nucleus (blue lines) are expected based on different nucleus durations. These are, however, not essential according to this approach. Thus, the relevant predictions of the uniformity hypothesis all refer to the distance from the beginning of the first syllable nucleus only:

U-P1 Peak timing is uniform across nuclear and non-nuclear conditions.
⇒ There is no main effect of nuclear condition.

U-P2 Peak timing is uniform across words with quantity 1 and quantity 2 vowels.
⇒ There is no main effect of quantity.

U-P3 ⇒ There is no interaction of quantity and nuclear condition.

For all of these predictions, only the contrast between two kinds of combined focus conditions and positions are relevant: those in which nuclear accents occur according to the nuclear accent hypothesis (subjects, verbs and objects in narrow focus and objects in broad focus) vs. those in which they are do not occur (unfocused given subjects, verbs and objects and subjects and verbs in broad focus). Also in the statistical modelling, this binary distinction was coded, since it was targeted at concrete hypothesis testing instead of providing the most parsimonious models. Note also that the nuclear accent hypothesis makes no prediction about whether or not an main effect of quantity should appear in addition to the interaction described in N-P2.

3.2.3 Results

Figure 3.7 and 3.8 illustrate segment durations and peak alignment in all experimental conditions: in sentence-initial, -medial and -final position (panels from top to bottom), in narrow focus, broad focus and unfocused condition (panels from left to right) and in quantity 1 and quantity 2 (bottom and top bars within each panel). Thereby, the nuclear conditions are marked by green bars and the non-nuclear ones by blue bars. Figure 3.7 shows the peak alignment found in the edited data set used in the present chapter (represented by black diamonds), while Figure 3.8 presents peak alignment in the data set evaluated in the previous chapter for comparison.[3] Figure 3.8 thus corresponds to Figure 2.16, albeit without indicating the average timing of low turning points.

Overall, the average timing of peaks in the edited data set was highly uniform across quantities and also very similar across focus conditions and sentence positions / grammatical functions, although peaks for non-nuclear conditions were realised later than for nuclear conditions in position 2 and even more in position 3 (see Figure 3.7). Though mean peak time was most often located during the first syllable nucleus for both data sets, it is noticeable at first sight that the alignment patterns found in the present, edited data set were more similar across the experimental conditions.

In the nuclear conditions, peaks were measured close to the beginning of the first syllable vowel nucleus. Visual inspection suggests that peaks in

[3] I will refer to these as the 'edited' and the 'original' data set in the following, although the 'original' data set used in the previous chapter was actually minimally edited (see Appendix C).

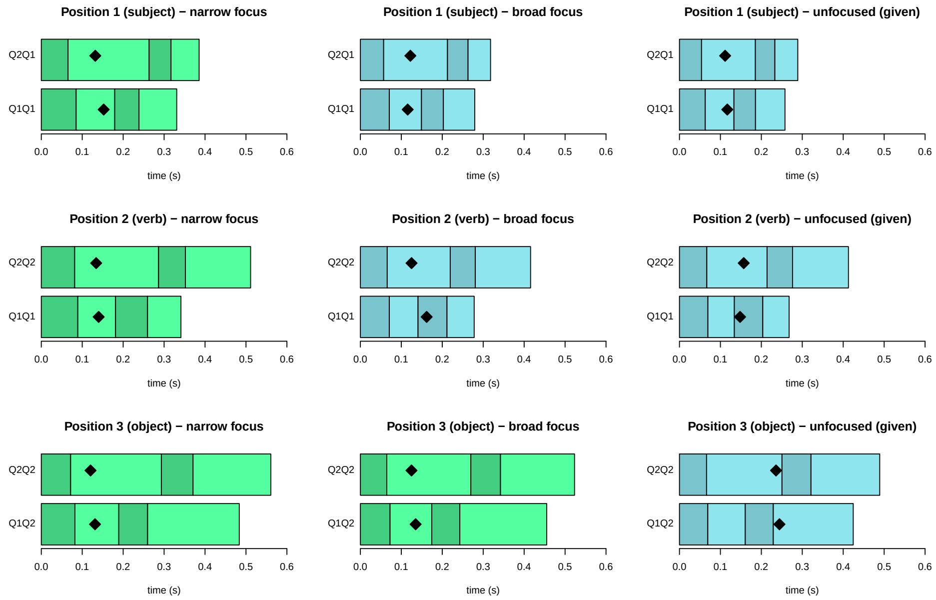

Figure 3.7: Average segment durations and times of f_0 turning points for all experimental conditions for the edited data set used in the present chapter.

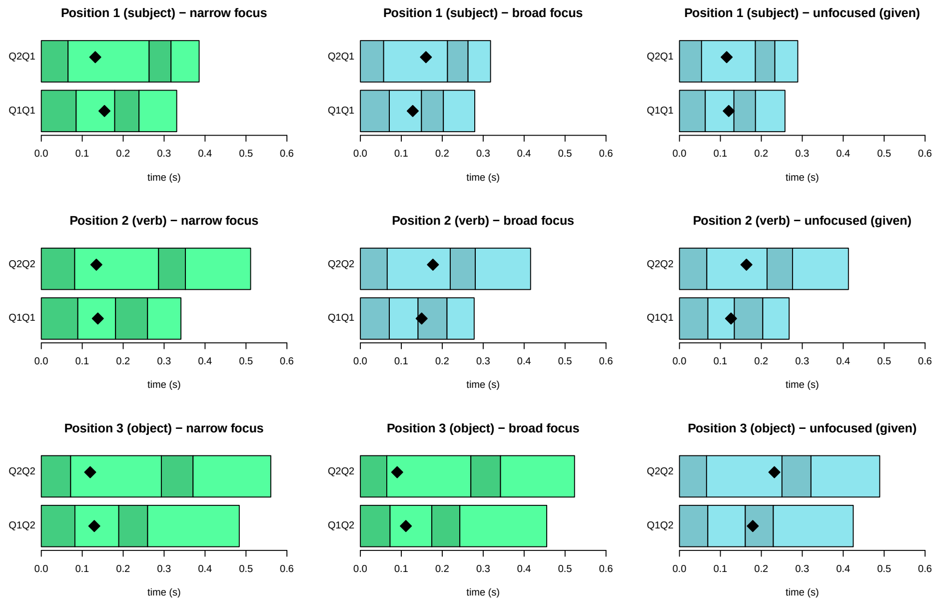

Figure 3.8: Average segment durations and times of f_0 turning points for all experimental conditions for the minimally edited original data set used in Chapter 2 (based on Figure 2.16).

quantity 2 were realised slightly earlier than those in quantity 1 as seen from the beginning of the word. The difference relative to the beginning of the nucleus, in contrast, seems negligible. Since relatively few data points were sorted out in the nuclear accent conditions (see section 3.2.1), the average timing was expectedly similar to what was observed for the original data set (see Figure 3.8). However, a smaller difference between the quantities than in the original data set was found for maxima of narrow focus subjects and especially for broad focus objects, which occurred on average a bit later in the edited set (probably due to removing 'fall from verb' contours).

In the non-nuclear conditions, the edited data set differed more clearly from the original one. Like for the nuclear conditions, peak timing in non-nuclear conditions was very similar between the two quantities in the edited data set according to Figure 3.7. The largest differences appeared for the verbs, which were least strongly represented in the data set—both overall (28% of the data) and within broad focus and unfocused conditions (29% and 23%, respectively). Also note that the quantities differed in two different ways for the non-nuclear verbs: In broad focus, quantity 2 showed slightly earlier peaks than quantity 1, whereas in unfocused condition, quantity 2 peaks were realised a bit later than those in quantity 1 words. In contrast, the timing was most uniform across quantities for the sentence-initial subjects, which made up the largest part of the data set (with 43%, 40% and 53% data points overall, in broad focus and unfocused, respectively).

A comparison of the figures further suggests that peaks in non-nuclear conditions were generally earlier in the edited than in the original data set, although verbs and objects (positions 2 and 3, respectively) in non-nuclear conditions constituted an exception. Subjective inspection of unfocused verbs suggested that many had a relatively small pitch range and that identified maxima partly stemmed from microprosody (recall that the lexical material included non-sonorants only in the second position, see section 2.2.1). However, Appendix F shows that different thresholds for determining perceptually relevant pitch movements did not yield earlier peaks for non-nuclear words in second (or third) position. At any rate, peaks of unfocused and broad focus verbs were only a little later in the currently used than in the original data set. In contrast, the 'backwards movement' due to editing was more substantial for the unfocused quantity 1 objects. Subjective inspection detected a substantive amount of final rises (also see Figure 2.14 in Chapter 2). The late timing for pitch maxima in unfocused sentence-final words will be examined more closely below, but notice here that, again, the difference between the quantities was smaller than for the original data set.

A crucial difference between Figures 3.7 and 3.8 is that the quantity-crossover was no longer present in the edited data set. In the original data set, quantity 2 words had later peaks than quantity 1 words in non-nuclear conditions (i.e. larger distances from the beginning of the nucleus), as described in section 2.4.4 and visible in Figure 3.8. The nuclear accent hypothesis described this crossover as an alignment of peaks to the end of the nucleus in non-nuclear conditions. It thus predicted that the pattern would persist in the edited data set. However, Figure 3.7 does not show later peaks for quantity 2 words than for quantity 1 words in non-nuclear conditions. Also, Appendix F illustrates that the picture remains largely the same and the crossover is still absent for two alternative editing criteria.

These findings are supported by the results of the linear mixed-effects models for the distance of peaks from the beginning and end of the first syllable nucleus (see Tables 3.2 and 3.3, respectively). Since the models were fit specifically to test the hypotheses introduced in section 3.2.2, they only included the two-level factors that the hypotheses refer to: quantity (whether a word's first syllable contains a quantity 1 or a quantity 2 vowel) and nuclear condition (whether or not a word was nuclear, i.e. final in a focus constituent).

Table 3.2: Coefficients of the best linear mixed-effects model of the peak distance from the beginning of the first syllable nucleus (in ms), random effects of nuclear condition by item, random by-subject interaction of quantity and nuclear condition.

	Estimate	Std. Error	t value
(Intercept)	101.4296	13.2259	7.6690
Quantity 2	0.5925	4.8567	0.1220
Nuclear condition	−44.9514	14.4156	−3.1182

Table 3.3: Coefficients of the best linear mixed-effects model of the peak distance from the end of the first syllable nucleus (in ms), random effects of nuclear condition by item, random by-subject interaction of quantity and nuclear condition.

	Estimate	Std. Error	t value
(Intercept)	34.8336	10.9325	3.1862
Quantity 2	−104.1065	6.1612	−16.8972
Nuclear condition	−78.2937	13.4190	−5.8345

Both models suggest that peaks were realised earlier in nuclear condition. The distance from the beginning of the first syllable nucleus was smaller, as the negative main effect in Table 3.2 indicates. A negative main effect of nuclear condition also figures in Table 3.3, which means that peaks of words in nuclear condition showed a *larger* distance from the end of the nucleus (recall that the measure was calculated as *time of peak (H) - time* *of beginning of nucleus*). The model in Table 3.3 additionally contains a negative main effect of quantity, reflecting that—as visible in Figure 3.7—peaks of quantity 2 words were consistently further away from the end of the first syllable nucleus. With respect to the the beginning of the nucleus, the quantities did not differ significantly (see Table 3.2).

Neither of the two models was improved by including an interaction between quantity and nuclear condition in the fixed effects. In the by-subject random effects, an interaction term was included and omitting this term significantly worsened the fit for both models. One can therefore ask whether the nuclear accent hypothesis and the interaction predicted as N-P2 in section 3.2.2 were not confirmed by the data set as a whole, but borne out for some participants. Appendix G, which inspects the by-subject adjustments and subsequently conducts sub-group models, indicates that this was not the case.

Returning to the special role of unfocused sentence-final words mentioned above, Figure 3.9 illustrates the distribution of both measures of alignment for nuclear and non-nuclear realisations of all items. Both panels show very narrow distributions for all items in nuclear condition and also for non-nuclear realisations of subjects and most verbs. In contrast, distributions are much more spread out for peaks of sentence-final non-nuclear (i.e. unfocused) items. Specifically, peaks occurring more than 300 ms after the beginning of the first syllable nucleus were measured much more often in this than in any other conditions. While these late maxima constituted 33% of the measured values for unfocused objects (i.e. 78 cases), they appeared on 7% (N = 6) of the broad focus objects, 1% (N = 3) of the unfocused verbs and 1% (N = 2) of the narrow focus subjects. In all other conditions, no such late maxima appeared at all.

Figures 3.10– 3.13 illustrate examples of late f_0 maxima on a subject and two objects, which are analysed as realisations of H_i tones here (marked in purple in the illustrations). Of the three cases of late maxima on verbs, two were partly obscured by non-modal voice quality, whereas the third one clearly showed a rise-fall contour spanning the whole VP (for an illustration, see Figure 3.1 above or Figures 2.3, 2.4 and 2.5 in Chapter 2). None of the contours showed a final rise on the verb and thus no reason to assume the presence of an H_i tone. In contrast, the late maximum on the subject example in Figure 3.10 (which is clearly visible as an outlier for item s2 in the upper panel of Figure 3.9) fits the term final rise very well. The narrow focus subject is followed by a pause, after which the speaker added the rest of the sentence—unfocused material mentioned in the preceding prompt question. This pause is analysed as the location of an i-phrase boundary right-edge aligned with the focused constituent in the current approach. Note also that the rise to the late maximum is preceded by a smaller pitch fall. It is interpreted as a realisation of the H_pL_p tones associated with the p-phrase. The other utterance with a late peak on a subject was very similar, likewise

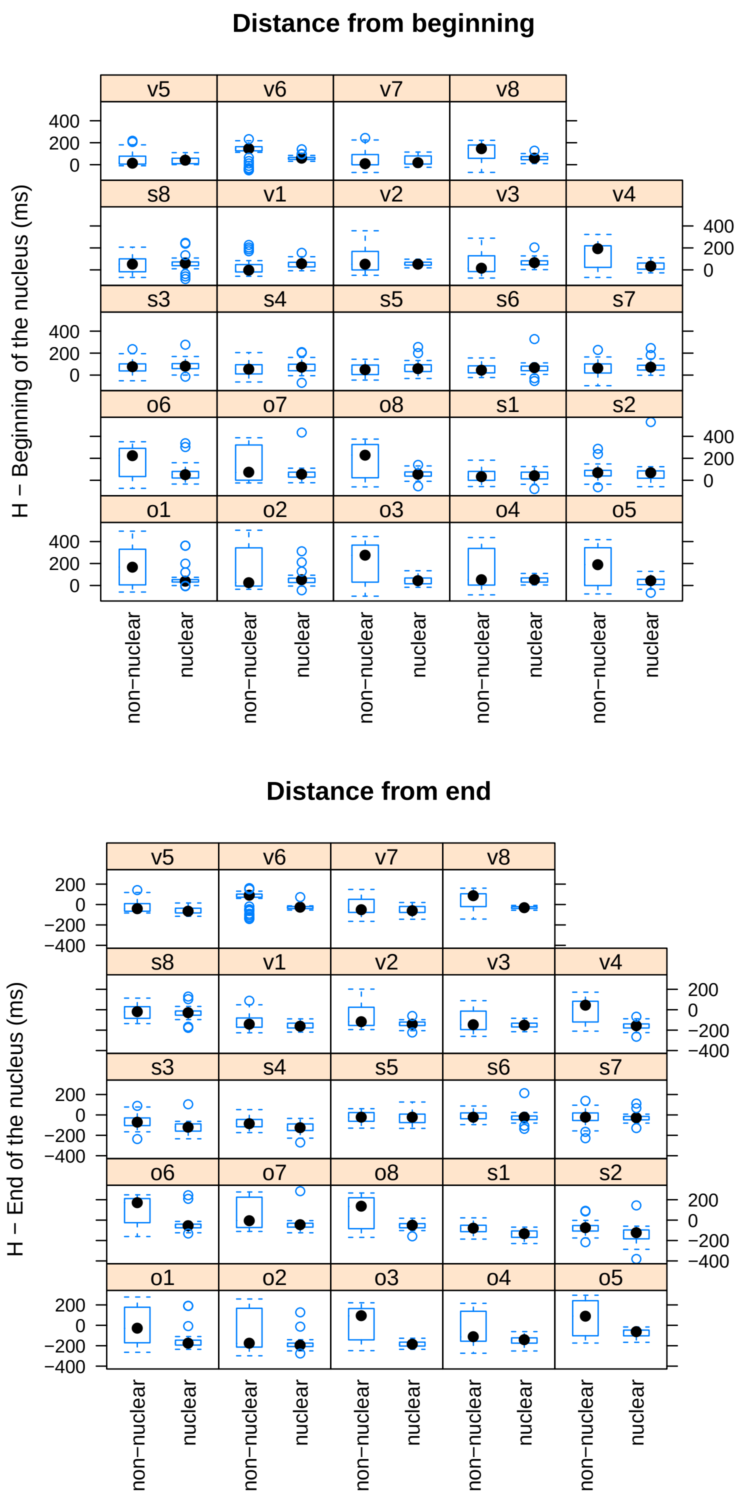

Figure 3.9: Boxplots for two measures of alignment by item, where items s1–s8 are subjects (position 1), v1–v8 are verbs (position 2) and o1–o8 are objects (position 3).

including a small fall preceding the rise and a pause before the speaker uttered the unfocused remainder of the sentence with a small pitch range.

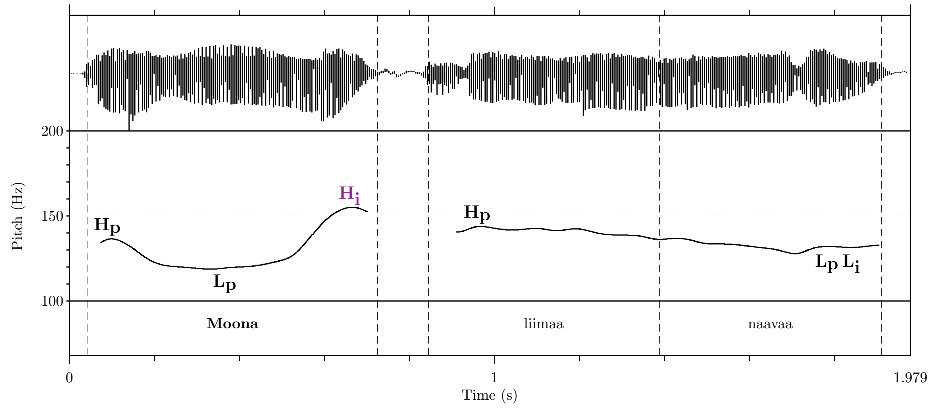

Figure 3.10: Example of final rise on the sentence-initial subject (in boldface). Sentence *Moona liimaa naavaa* 'Moona glues lichen', information focus on the subject.

Contours with late f_0 maxima on the sentence-final objects were more varied and often (partly) obscured by non-modal voice quality. Figure 3.11 depicts a broad focus realisation where a high rise follows a rise-fall contour spanning the VP, with its peak on the verb (the end of the fall and the beginning of the final rise are hidden by creaky voice). In contrast, the final rise on the unfocused object in Figure 3.12 ends at a comparatively much lower level, which does not exceed that of the preceding peaks in the sentence (also, the narrow focus verb carries a H_pL_p tones of its own). Figure 3.13 illustrates another relatively low final rise that is more gradual, starting from an earlier preceding L_p of an HL realisation spanning the whole VP. All these final rises are analysed here as realisations of an H_i tone.

As a result of these frequent late peaks in unfocused sentence-final words, distributions of peak timing for non-nuclear position 3 items were broader in Figure 3.9. Also the median values, marked by black dots, showed a much larger difference between nuclear and non-nuclear conditions for sentence-final objects than in any other position. This ties in with the exceptionally late mean peak timing of non-nuclear objects observed for Figure 3.7.

Therefore, two additional subset models were fit to the data after excluding measurements from unfocused words in sentence-final position. The results were quite different from those for the complete data set. For the subset model of the distance of the peak from the beginning of the nucleus, no significant effect of quantity or nuclear condition was found (see Table 3.4). Also adding an interaction between the two factors did not improve the fit. In contrast, the model of the distance from the end of the nucleus in Table 3.5

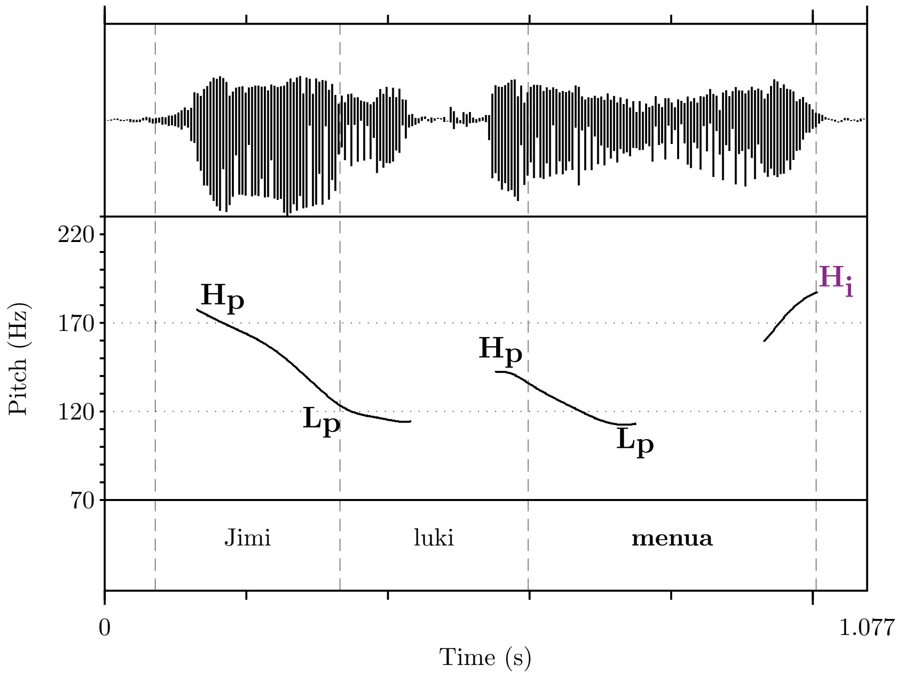

Figure 3.11: Example of high final rise on the sentence-final object (in boldface). Sentence *Jimi luki menua* ‘Jimi read the menu’, broad focus.

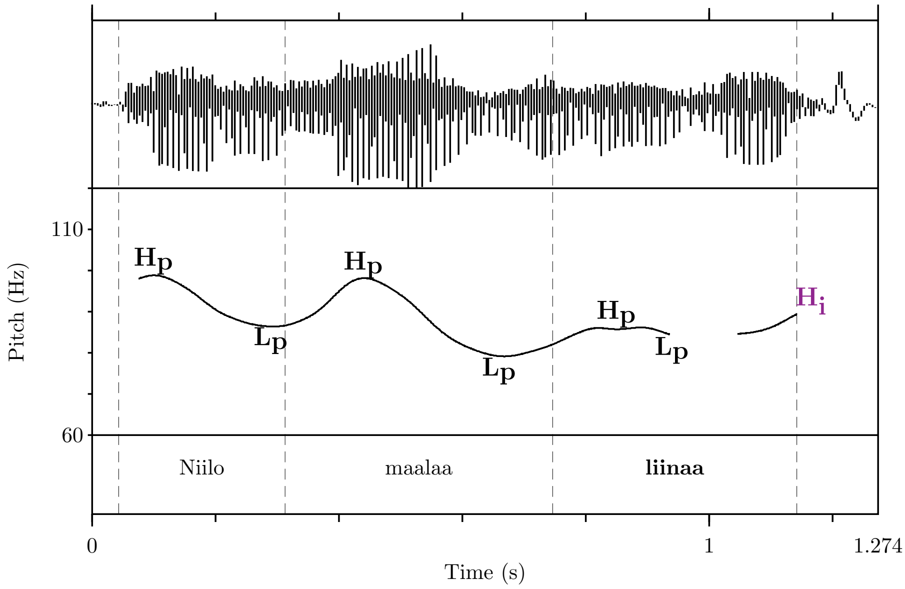

Figure 3.12: Example of low final rise on the sentence-final object (in boldface). Sentence *Niilo maalaa liinaa* 'Niilo paints a cloth', contrastive focus on the verb.

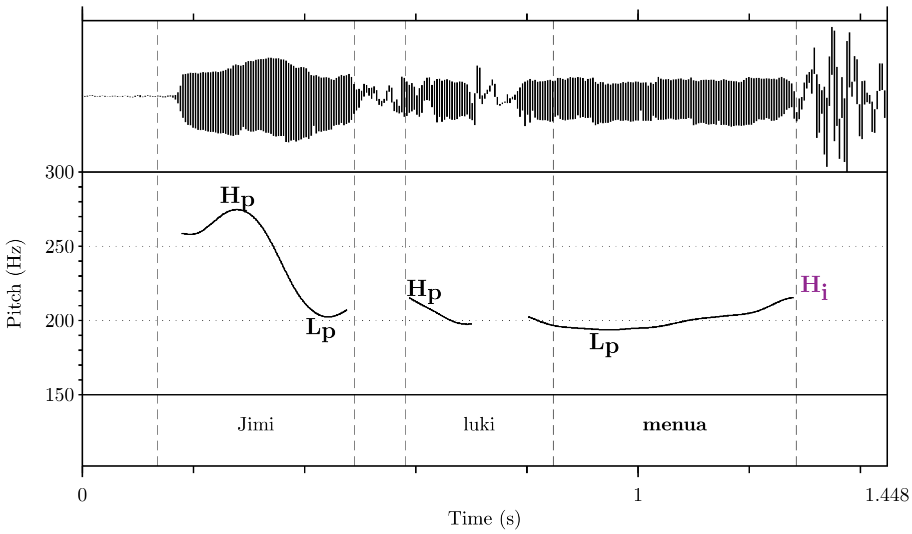

Figure 3.13: Example of low final rise on the sentence-final object (in boldface). Sentence *Jimi luki menua* 'Jimi read the menu', information focus on the subject.

showed negative main effects of both factors and an additional interaction. Peaks were further away from the end of the nucleus in quantity 2 words than in quantity 1 words and further away in nuclear conditions than in non-nuclear conditions. Peaks of quantity 2 words in nuclear condition were also further away, i.e. earlier relative to the end of the nucleus, than peaks in all other condition by a margin that exceeded the combined effects of both factors (see the interaction in Table 3.5).

Table 3.4: Coefficients of the linear mixed-effects model of the peak distance from the beginning of the first syllable nucleus (in ms) for data set without unfocused sentence-final words, random effects of nuclear condition by subject and of nuclear condition by item.

	Estimate	Std. Error	t value
(Intercept)	68.1434	8.9885	7.5811
Quantity 2	0.5747	4.7015	0.1222
Nuclear condition	−11.7937	9.4084	−1.2535

Table 3.5: Coefficients of the linear mixed-effects model of the peak distance from the end of the first syllable nucleus (in ms) for data set without unfocused sentence-final words, random effects of quantity and nuclear condition by subject and of nuclear condition by item.

	Estimate	Std. Error	t value
(Intercept)	−3.9940	11.9721	−0.3336
Quantity 2	−72.9371	16.6362	−4.3842
Nuclear condition	−37.0934	12.6226	−2.9386
Quantity 2 : Nuclear condition	−37.0102	17.1666	−2.1559

To complete the picture of peak alignment in the edited data set, models were also fit using the basic experimental factors as predictors. That is, instead of only distinguishing between conditions in which the respective word was final in a focus constituent (nuclear) and those in which it was not, as in the preceding models, they employed all levels of the three factors position, focus condition and quantity. For both measures of alignment, two models were fit: one to the complete data set and one after excluding unfocused sentence-final words. Without these, the model of the distance from the beginning of the first syllable nucleus showed no significant effects at all (see Table 3.6). It only indicated that peaks in position 3 were overall marginally earlier than those in other positions (see the t-value of −1.9912). The best model also did not include any interactions.

Table 3.6: Coefficients of the linear mixed-effects model of the peak distance from the beginning of the first syllable nucleus (in ms) for data set without unfocused sentence-final words, considering all experimental factors. Random effects of item and of focus condition and position by subject. CF indicates contrastive focus, IF information focus on the narrow focus constituent.

	Estimate	Std. Error	t value
(Intercept)	71.7637	7.8573	9.1333
Quantity 2	−0.4074	4.0184	−0.1014
Narrow focus (CF)	−2.2026	8.2589	−0.2667
Narrow focus (IF)	−12.3614	8.1182	−1.5227
Unfocused (CF)	2.7462	8.8184	0.3114
Unfocused (IF)	6.9723	8.7284	0.7988
Position 2	−11.4768	9.7947	−1.1717
Position 3	−14.4511	7.2574	−1.9912

The corresponding model for the complete data set likewise did not exhibit significant main effects (see Table 3.7). Its interactions did, however, indicate that differences between the focus conditions were stronger sentence-medially and -finally than for the sentence-initial intercept, although this was not always significant. These interactions generally pointed to narrow focus words having earlier and unfocused words later peaks in positions 2 and 3. However, while the interactions mean that effects of focus conditions were significantly larger in these positions, they do not signal that they were significant. Fitting further subset models showed that peak timing relative to the beginning of the nucleus did not differ significantly for verbs in different focus conditions. In contrast, the subset model for words in position 3 confirmed that peaks in this position were later in unfocused condition, in line with the interactions with the largest estimated effects at the bottom of Table 3.7.[4]

[4]For verbs (position 2), the model of peak distance from the beginning of the first syllable nucleus had random effects of focus condition by item and of focus condition by subject. It showed no significant effects and was not improved by adding an interaction:

	Estimate	Std. Error	t value
(Intercept)	73.3019	14.1545	5.1787
Quantity 2	−1.2129	7.2495	−0.1673
Narrow focus (CF)	−19.5645	15.5980	−1.2543
Narrow focus (IF)	−22.5922	15.1836	−1.4879
Unfocused (CF)	1.6547	18.4920	0.0895
Unfocused (IF)	18.9940	16.5829	1.1454

For objects (position 3), the respective model had random effect of item and of quantity and focus condition by subject. It was not improved by adding an interaction term. It ev-

Table 3.7: Coefficients of the linear mixed-effects model of the peak distance from the beginning of the first syllable nucleus (in ms) for complete data set, considering all experimental factors. Random effects of item and of focus condition and position by subject.

	Estimate	Std. Error	t value
(Intercept)	55.9180	9.8479	5.6782
Quantity 2	2.5050	5.8271	0.4299
Narrow focus (CF)	14.9274	10.4570	1.4275
Narrow focus (IF)	4.1721	10.9999	0.3793
Unfocused (CF)	−1.2382	10.3294	−0.1199
Unfocused (IF)	−2.7796	10.6981	−0.2598
Position 2	15.2514	15.1474	1.0069
Position 3	7.7405	17.9622	0.4309
Narrow focus (CF) : Position 2	−34.2798	13.5396	−2.5318
Narrow focus (IF) : Position 2	−26.8453	13.5786	−1.9770
Unfocused (CF) : Position 2	6.2033	13.2557	0.4680
Unfocused (IF) : Position 2	27.2121	13.2899	2.0476
Narrow focus (CF) : Position 3	−22.2385	13.3668	−1.6637
Narrow focus (IF) : Position 3	−28.5088	13.4217	−2.1241
Unfocused (CF) : Position 3	110.8368	13.2612	8.3580
Unfocused (IF) : Position 3	92.8196	12.9035	7.1933

Seen from the end of the first syllable nucleus, a rather different picture emerged. When excluding unfocused sentence-final words, no interactions arose, but the model included significant main effects of all factors (see Table 3.8). It demonstrated that peaks were relatively earlier in quantity 2 words than in quantity 1 words and earlier in narrow focus than for the broad focus intercept. It also pointed to later peaks for unfocused words, although that effect was only marginal for contrastive focus. Lastly, the model indicated marginally earlier peaks for sentence-medial and significantly earlier peaks for sentence-final words compared to the position 1 intercept.

idences significantly later peaks for unfocused words. Otherwise, it displays no significant effects:

	Estimate	Std. Error	t value
(Intercept)	58.0674	11.1845	5.1918
Quantity 2	4.6658	10.6192	0.4394
Narrow focus (CF)	−3.1770	11.9490	−0.2659
Narrow focus (IF)	−20.8418	12.1389	−1.7169
Unfocused (CF)	102.1492	28.6754	3.5623
Unfocused (IF)	86.6175	28.9137	2.9957

Table 3.8: Coefficients of the linear mixed-effects model of the peak distance from the end of the first syllable nucleus (in ms) for data set without unfocused sentence-final words, considering all experimental factors. Random effects of focus condition by item and of focus condition, quantity and position by subject.

	Estimate	Std. Error	t value
(Intercept)	5.9990	8.6358	0.6947
Quantity 2	−111.6812	5.5637	−20.0732
Narrow focus (CF)	−27.7187	9.3367	−2.9688
Narrow focus (IF)	−31.3896	8.4515	−3.7141
Unfocused (CF)	17.5736	9.2020	1.9098
Unfocused (IF)	22.1767	9.3764	2.3652
Position 2	−17.8654	9.1617	−1.9500
Position 3	−36.2170	7.7841	−4.6527

Including peaks of unfocused position 3 words, only the main effect of quantity remained significant, while peaks of words in narrow information focus and in position 3 were estimated to be marginally earlier relative to the end of the nucleus (see Table 3.9). Additionally, this model of the complete data set encompassed interactions between the factors focus condition and position. These indicated that the effects of both types of narrow focus were stronger in position 2. Two further interactions again corroborated the exceptionally late peaks in unfocused sentence-final words.

3.2.4 Discussion

In the edited data set analysed in the current chapter, peak alignment was found to be fairly stable across experimental conditions, especially when measured relative to the beginning of the first syllable nucleus. Changes in first syllable vowel quantity resulted in varying distances from the end of the nucleus, but seen from its beginning, quantity did not have any effect. And while the overall models pointed to earlier peaks in nuclear condition, models excluding a specific type of non-nuclear words—unfocused sentence-final objects—again only showed earlier nuclear peaks relative to the end of the nucleus. Moreover, also a more detailed analysis of this reduced data set, distinguishing the two quantities, three positions, and five different focus conditions induced in the experiment, did not yield any significant differences in timing relative to the beginning of the nucleus, but only showed differences relative to the nucleus end.

Evaluating the predictions derived from the nuclear accent hypothesis in section 3.2.2, it would be difficult to count these findings as a reliable confir-

Table 3.9: Coefficients of the linear mixed-effects model of the peak distance from the end of the first syllable nucleus (in ms) for complete data set, considering all experimental factors. Random effects of focus condition by item and of focus condition, quantity and position by subject.

	Estimate	Std. Error	t value
(Intercept)	−6.7477	11.9932	−0.5626
Quantity 2	−109.3205	6.9433	−15.7447
Narrow focus (CF)	−16.4792	12.5248	−1.3157
Narrow focus (IF)	−19.4058	11.7755	−1.6480
Unfocused (CF)	15.2117	12.6894	1.1988
Unfocused (IF)	13.6634	14.3779	0.9503
Position 2	17.1565	17.1580	0.9999
Position 3	−31.6876	19.2960	−1.6422
Narrow focus (CF) : Position 2	−35.7654	16.6994	−2.1417
Narrow focus (IF) : Position 2	−34.4045	15.4358	−2.2289
Unfocused (CF) : Position 2	−1.3332	16.4746	−0.0809
Unfocused (IF) : Position 2	21.4461	18.8321	1.1388
Narrow focus (CF) : Position 3	−2.3446	16.5185	−0.1419
Narrow focus (IF) : Position 3	−6.6895	15.2536	−0.4386
Unfocused (CF) : Position 3	114.3271	16.4600	6.9457
Unfocused (IF) : Position 3	96.6619	18.5657	5.2065

mation of only nuclear peaks being aligned to the beginning of the nucleus and thus earlier (as was stipulated in N-P1). Instead of showing a division between earlier nuclear and later non-nuclear peaks, the data suggested that a seizable number of late peaks only occurred in a specific condition. Late peaks can thus not be taken as characteristic of non-nuclearity per se. In turn, defining a class of nuclear peaks by its early timing is moot. The second prediction of the nuclear accent hypothesis, that of an interaction between the factors (N-P2), was only confirmed for the data set without the unfocused sentence-final words and, importantly, only for the distance of the peak from the end of the syllable nucleus. Interestingly, the mean peak timing in the edited data set corresponded even less than the means from the original set to the the schematisation of the nuclear accent hypothesis in Figure 3.6, as the comparison between Figures 3.7 and 3.8 revealed. All in all, the present findings failed to provide sufficient evidence for the nuclear accent hypothesis.

The uniformity hypothesis, in contrast, was in better agreement with the data. For the distance from the beginning of the nucleus, there was no effect of quantity and no interaction between quantity and nuclear condition, as predicted in U-P2 and U-P3, respectively. Furthermore, the models of the distance from the end of the nucleus were also in line with the uniformity hypothesis. Larger distances or relatively 'earlier' peaks were found in conditions that induced longer durations, i.e. in quantity 2 and in nuclear conditions, especially in narrow focus (see again section 2.4.3). This fits in well with the assumption that peak alignment was fixed relative to the beginning of the nucleus, so that that the distance to the nucleus end was determined by nucleus duration (i.e. the distance between beginning and end of the nucleus). The beginning of the nucleus thus appears to be a suitable landmark relative to which peak alignment can be described consistently, whereas findings relative to the end of the nucleus were redundant.

In contrast to U-P2 and U-P3, the first prediction from the uniformity hypothesis was not borne out by the complete data set. U-P1, claiming the absence of an effect of nuclear condition on alignment relative to the beginning of the nucleus, was only confirmed for the model in Table 3.4, which did not consider unfocused sentence-final words. However, it will be argued here that it is in fact appropriate to disregard f_0 maxima of unfocused sentence-final words from the current analyses of peak timing. The basis for this argument is that, according to the interpretation suggested here, many maxima of unfocused objects were realisations of H_i tones. Obviously, the timing of the H_i tones should not be considered in an investigation of the alignment of H_p tones.[5]

[5]Strictly speaking, a more precise approach than excluding all maxima of unfocused sentence-final objects would have been to only dismiss realisations of H_i tones (including those few in other experimental conditions). For this, the data would however need to be manually edited.

To explain this in a bit more detail: Section 3.2.3 showed that in unfocused objects, measurements showed larger variation and later averages than in all other conditions. The section further argued that this was due to the frequent occurrence of final rises. When the end point of a final rise was the highest f_0 value measured for a word, it was included in the edited data set for the analysis of peak alignment. This seems inappropriate, because late maxima did not emerge from the inspection of the individual utterances as a form of peak realisation that contrasts with the early peaks prevalent in the majority of experimental conditions. The data suggested that the largest group of late maxima, termed final rises here, should not in fact be considered peaks. Instead of contrasting with early peaks, they often co-occurred with them. The given examples of final rises showed that these were preceded by pitch falls—often on the same word, as in Figures 3.10 and 3.12. The analysis suggested here is that these falls are realisations of H_pL_p tones, while the final rise is due to a following H tone associated with a higher ranking phrase: H_i. Figures 3.11 and 3.13 show cases where H_p was already realised on the preceding word, so that L_p and H_i together formed the final rise.

Other possible interpretations of the present findings should be mentioned. Of course, it can be discussed how the prosodic domains to which tones are ascribed here should be termed. More interestingly, one can ask whether all of the tones should be viewed as associated with a phrasal unit—i.e. as phrasal tones—as is done here. Specifically, this might be contested for H_pL_p. Empirically, this question is very difficult to decide. The results presented above are taken as confirmation of the uniformity hypothesis. The uniformity hypothesis as proposed here states that peak alignment is uniform across quantities, positions and information structural conditions. Thus, it does not recognise a categorical difference between nuclear and non-nuclear accents. It refuses the idea that there are different kinds of accents differentiating quantities, positions or focus conditions at all.[6] However, the hypothesis is compatible with either the assumption of uniform accents or another uniform combination of tonal targets, e.g. phrase tones.

While the second option is adopted here, previous literature on Finnish prosody has prevalently espoused the first approach (e.g. Välimaa-Blum, 1988, 1993; Iivonen, 1998; Suomi et al., 2008). Additional alternatives can also be entertained: While usually both of the p-phrase tones are interpreted as accentual tones instead (see Välimaa-Blum's 1993 L+H* and Suomi et al.'s 2008 LHL), one could claim that only the L_p marks the (boundary of a) p-phrase, which is combined with a uniform H* accent. This would allow to account for the possibility of other, contrasting accents that the present

[6]This refers to accents in the AM-sense, i.e. different tonal specifications associated with a prominent prosodic position. The current chapter only investigated peak timing, which Chapter 2 identified as the most likely difference between different types of accents in the current materials, should they exist.

materials failed to uncover (possibly the hinting delayed peak described by Iivonen, 1987, 1998). In the absence of such other accent types, it is not easy to distinguish an H*L or H*L_p account from the H_pL_p analysis advocated here. In particular, it is not completely clear whether the three options would predict a different alignment of the L tone relative to the H.

Finally, also the present observations of final rises can be related to existing literature. Ogden and Routarinne (2005) investigate what they call rising terminal (RT) intonation contours in conversations of teenage girls from Helsinki, using a combination of phonetic measurements and Conversation Analysis. They describe RTs as "a rising intonation contour which starts on the last accentually prominent syllable of the TCU [turn-constructional unit, AA]" (Ogden and Routarinne, 2005, p. 163), which "ends mid-high to high in the speaker's range" (ibid.). Citing Välimaa-Blum (1993), they refer to falls such as H*L or level tones followed by a final H%. For these contours, which are very frequent in their materials, Ogden and Routarinne (2005) identify two main functions: They mark a possibility for the listener to respond with a reception signal and they indicate that the speaker will continue talking (Ogden and Routarinne, 2005, p. 167). They could thus be referred to as a continuation rises.

The more general term 'final rise' was adopted here to mark that it is not clear whether these are rising terminals as described by Ogden and Routarinne (2005) or not, since it is difficult to compare the current data to (pictures of pitch tracks and transcriptions from) their conversational material. Also, one might ask why the rises as described by Ogden and Routarinne (2005) should appear in laboratory-recorded data at all. There are at least two possible reasons: First, these laboratory recordings might have had an interactional aspect after all and second, final rises might just generally be a part of an individual's speaking style.

As to the first reason, the function of projecting more talk by the speaker that Ogden and Routarinne (2005) found for RTs is likely shared at least for the two final rises on subjects (cf. Figure 3.10). The two subjects with final rises would sound odd without the continuation of the sentence after the pause, while only the same narrow focus subjects with a falling contour would in themselves be perfectly well-formed answers to the preceding Wh-questions. Indeed, these one-word answers would be more natural than additionally repeating the contextually given materials like participants in the current experiment were asked to do. In this vein, it is interesting that final rises (at least the ones that resulted in the word's pitch maximum) predominantly occurred on post-focal and previously mentioned (given) constituents that would normally be elided. Since I as the experimenter was in the same room and sat fairly close to the participants, it is thus not unlikely that they employed a contour that functions to elicit a reception signal to ascertain that their production was as desired (every experimenter knows that participants will frequently ask if they got things right and passed the test, even if they are explicitly told that experiments are about testing theories, not participants).

Regarding the second issue, the participants of Production Experiment I belong to the group of Finnish speakers known to produce the final rises described by Ogden and Routarinne (2005): young people from Helsinki. In their discussion of rising intonation in Finnish, Suomi et al. state that according "to Routarinne (2003), rising intonation is becoming more and more common in the speech of young women particularly in the Finnish capital area (Helsinki and its surroundings)" (Suomi et al., 2008, p. 119). They further remark: "Any speaker of Finnish, a native or second-language speaker, has certainly observed speech situations involving clearly rising intonation in (colloquial) Finnish" (Suomi et al., 2008, p. 117)—and I am no exception. I have frequently encountered these specific final rises when interacting with native speakers or listening to them interact. Based on this, I have the strong impression that this is a feature that characterises the speech of some individuals, while largely pertaining to a certain sociological group: young people in the Helsinki region (I have heard these rises employed also by some young men very frequently). Native speakers have confirmed to me that these final rises are common among certain young speakers from the Helsinki area, but seem alien to Finns with different backgrounds. If a particular intonational feature is characteristic of an individual's speaking style, there is no a priori reason why that speaker should not employ this feature in a recording session. Also recall that data from speaker s01 was discarded completely because she used final rises very persistently (see section 2.2.2).

Clearly, more research is needed to confirm that there are functionally different types of rises, as well as to establish the respective contexts in which they occur. Impressionistically, the utterance illustrated in Figure 3.11 sounds like the specific type of final rises that I am fairly certain are the ones studied by Ogden and Routarinne (2005), whereas the other examples of final rises depicted above do not. Presently, it is not clear whether young-Helsinki-dweller's final rises are based on a different phonological specification of tonal targets than other types of final rises, presuming the distinction can be confirmed in empirical research. In the absence of any decisive evidence, no difference is made here and all final rises were annotated with H_i tones above. Should a phonological distinction prove appropriate, one could posit that there are three possibilities for the tonal specification of a Finnish i-phrase: L_i, H_i, and the lack of a tonal specification (for additional clarity, this could be notated as 0_i, following Grabe's 1998 notation 0%). The first one could be used to describe the most frequent realisation, namely a focus phrase ending in a pitch fall (close) to the bottom of the speaker's range, often masked by non-modal voice quality (see Välimaa-Blum's 1993 L%). The second notation would be used for high final rises like those described by Ogden and Routarinne (2005) (corresponding to their H%), while the absence of a tonal specification (or 0_i) could signify cases where pitch neither drops nor rises to a high level (possibly what Välimaa-Blum 1993 accounted for as H%, see p. 87: "the continuation tune [...] is almost horizontal toward the boundary which I propose is H%"). This proposal would need to be verified by further empirical research and is therefore not further pursued here.

In contrast, this chapter's findings have provided ample evidence for the existence of a distinction between H_i and L_i tones.

3.3 A pilot study on the perception of peak height and peak timing

This section presents the results of a small perception study—a pilot for an alternative to the experiment discussed in Chapter 5—investigating the impact of f_0 height and timing of pitch peaks. Starting from the finding of extended pitch falls on narrow focus constituents in Chapter 2, the aim of the study was to find out whether earlier peaks and thus longer pitch falls could function as markers of narrow focus per se. This would constitute strong evidence for the nuclear accent hypothesis claiming a difference between a nuclear falling accent and a non-nuclear rising accent (preliminarily designated H*L vs. LH*). However, the study did not turn up any evidence for such a distinction. Participants identified items with higher peaks as narrowly focused more often, but did not show a reaction to differences in peak timing.

3.3.1 Data and analysis

Four university-educated native speakers of Finnish with otherwise slightly different backgrounds participated in the experiment in May 2011 (two women and two men; an undergraduate student in her twenties, a graduate student in her thirties and two researchers around 50; three of them from the Helsinki area and one from the Oulu region). Table 3.10 gives the lexical materials used in the experiment. The participants' task was to judge on a 1 to 5 scale how likely a sentence they heard answered a question presented in written form on a screen, with 1 signifying the judgement as least likely and 5 as most likely. The written prompt read 'Was the question X?', with X being the respective question in Table 3.10, and the response buttons below it were labelled 'surely not' (=1), 'unlikely' (=2), 'maybe' (=3), 'likely' (=4) and 'surely' (=5).

In each sentence, there was a target word (in bold print in Table 3.10) and the corresponding question induced narrow information focus on this word. The timing and height of the target word's f_0 maximum were manipulated to find out whether these changes would cause the participants to perceive it as narrowly focused. Timing and f_0 were always manipulated in three equidistant steps, with the second step corresponding to the original

Table 3.10: Stimuli used in the pilot perception experiment on f_0 height and timing of pitch peaks. Time steps are given in ms, f_0 steps in st.

	Question	Answer	Time step	F_0 step
1	*Kenelle Veera antoi kuitin mekkosta?* 'To whom did Veera give the receipt for the dress?'	*Veera antoi* ***myyjälle*** *kuitin mekkosta.* 'Veera gave **the sales clerk** the receipt for the dress.'	100	3
2	*Mitä Laura leipoi maanantaina pikkujoulujuhliin?* 'What did Laura bake on Monday for the pre-Christmas party?'	*Laura leipoi maanantaina* ***kakun*** *pikkujoulujuhliin.* 'Laura baked **a cake** on Monday for the pre-Christmas party.'	80 60	4 4
3	*Mitä mieltä olet siitä, että odottaa Jonna vielä bussia?* 'What is your opinion about Jonna still waiting for the bus?'	*Jonna* ***valitettavasti*** *odottaa vielä bussia.* 'Jonna **unfortunately** still waits for the bus.'	100	2
4	*Kuka ostaa heinäkuussa veneen?* 'Who will buy a boat in July?'	***Marianna*** *ostaa heinäkuussa veneen.* '**Marianna** will buy a boat in July.'	60	4

peak time and f_0 measured in the source sound. For the other words in the item sentences, the manipulated contour was a close approximation of the original pitch. Most of the source sounds for the manipulations were fillers from Production Experiment I realised by different participants (see Chapter 2), but item 1 was produced by one of the pilot study's participants. Actual step sizes differed between the items, as given in the table. Time steps were adjusted to avoid peaks outside the target words, while f_0 step sizes were chosen to suit the individual pitch range of the respective source sound speakers. Item 1 was used for two sets of manipulations, once with and once without an added low tone at the beginning of the target word—corresponding to L1 in the measurements performed on the production experiments in this thesis. Also, two versions of item 2 by different speakers were used for separate sets of manipulations, one of them with a slightly longer pause between the target word and the following one. As indicated in Table 3.10, different steps were used for both versions. Item

4 was only presented to two of the four participants. For two participants, the question in item 1 was *Kenelle antoi Veera kuitin mekkosta?*, with the different word order implying a slightly different information structure for the question, but still narrow focus on the target word in the answer. Note also that the word order in the question for item 3 is slightly odd and should read ... *että Jonna odottaa* instead of ... *että odottaa Jonna*.

For the pilot perception study, the manipulated sounds were interspersed with nine filler question-answer pairs. While the participants encountered all 3 x 3 = 9 variants of the target items, the filler items were only presented once, resulting in 54 items per session for the first two participants and 63 items for the other two. Presentation order was pseudo-randomised and differed for each participant. Pitch manipulation and response data collection were conducted using Praat, specifically manipulation objects plus PSOLA re-synthesis and ExperimentMFC, respectively (Boersma and Weenink, 2010). The resulting response data were evaluated with linear mixed-effects models as implemented in R (R Development Core Team, 2010; Baayen, 2008). The two versions of item 1 and item 2 were each treated as separate items in the modelling. For one of the participants, the first 31 trials were disregarded, since the participant took very long to give responses and asked for a clarification of the instructions, so that the overall number of evaluated responses was 175.

3.3.2 Results

As Figure 3.14 shows, participants clearly judged the target items with higher peaks as more likely to be narrowly focused. That is, when a sentence included a target word with the third and highest f_0 step, it was rated as significantly more congruent with a question inducing narrow focus on the target word, as the model in Table 3.11 supports. The model also suggests that the effect of lowered pitch was likewise significant, although Figure 3.14 illustrates that the effect was smaller. Thus, targets with manipulated pitch peaks higher than the original value were more likely to be perceived as narrowly focused while those with lower peaks were less likely to be perceived as narrowly focused.

Table 3.11: Coefficients of the best mixed-effects model of participant response, with random by-item effects of f_0 step and random by-subject slopes.

	Estimate	Std. Error	t value
(Intercept): original f_0	1.8071	0.4165	4.3388
F_0 level 1: lower f_0	−0.3156	0.1277	−2.4717
F_0 level 3: higher f_0	0.7768	0.2739	2.8357

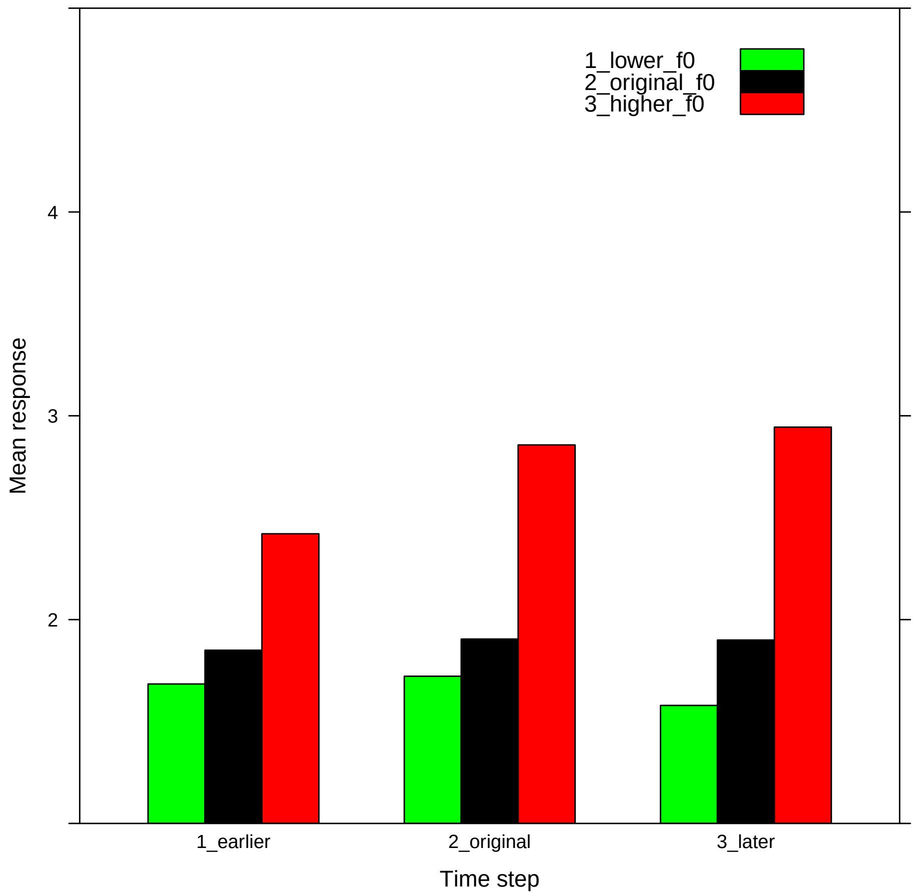

Figure 3.14: Mean responses to stimuli with manipulated peak timing and f_0 on a scale from 1 $\sim$ 'surely not narrowly focused' to 5 $\sim$ 'surely narrowly focused'.

In contrast, there was no indication of an effect of manipulating target peak timing. Figure 3.14 suggests that judgements as narrowly focused were overall slightly less frequent for targets with earlier peaks, but adding time step as a factor to the model in Table 3.11 did not enhance model fit. Adding an interaction between f_0 step and time step did not improve the model, either, as an ANOVA comparison indicated.[7]

3.3.3 Discussion

The results of this pilot perception study corroborated the finding that narrow focus constituents in Finnish are realised with a larger pitch range and higher pitch peaks. Thus, they fit in well with the production studies reported in Chapter 2 and in the previous literature (see Välimaa-Blum, 1988, 1993; Mixdorff et al., 2002; Vainio and Järvikivi, 2007). More directly, they also confirm Vainio and Järvikivi's (2006) perception experiments showing that changes in the pitch contour of a sentence affect how Finnish listeners interpret the information structure of this sentence.

In that light, the absence of evidence for an effect of the previously not investigated factor peak timing seems compelling. In spite of the small size of the data set, a meaningful, consistent and statistically significant effect emerged for the f_0 height manipulations, making the lack of such an effect for the manipulations of timing appear all the more telling. Therefore, the data presented in the current section are taken as further tentative evidence for the uniformity of Finnish peak timing and against a phonological distinction between accents as proposed by the nuclear accent hypothesis.

Of course, at least two counter-arguments can still be made. First, it is possible that the size of the chosen time steps was simply inadequate (for the respective stimuli) and thus failed to induce a reliable response. While this objection cannot be invalidated completely, it is worth noting that the step sizes were quite large and certainly audibly salient to my (albeit non-native) ears. Second, the induced timing variations could be both perceivable and phonologically meaningful to Finnish listeners, but marking something else than the presence vs. absence of narrow focus on the respective constituent. Indeed, the results of this pilot strongly suggest that the nuclear accent hypothesis should be rejected, but the possibility of a different phonological

[7] Table 3.11 models the manipulation as factorial, i.e. consisting of three distinct steps. The best alternative model treating peak f_0 as a numeric vector (with the three values -1 $\sim$ 'lower f_0', 0 $\sim$ 'original f_0' and 1 $\sim$ 'higher f_0') is given below. It corresponded to Table 3.11, showing the same main effects and random structure. It likewise indicated a significant effect of the f_0 manipulations, while peak time manipulation—as a main effect or interaction—did not improve model fit.

	Estimate	Std. Error	t value
(Intercept)	1.9581	0.4127	4.7447
F_0	0.5534	0.1498	3.6933

distinction in peak timing cannot be dismissed. The present data thus only constitute a partial argument for phonological uniformity of peak timing. The pilot study was constructed specifically with the idea of H*L focus or nuclear accents in mind. The study failed to provide evidence supporting the postulation of this accent or the nuclear accent hypothesis. Instead, the results in line with the uniformity analysis an analysis in terms of H_pL_p tones.

3.4 Stress and intensity

This section presents a post-hoc analysis of intensity in the data from Production Experiment I. Its main purpose is to establish whether intensity is a correlate of primary stress in Finnish. While the materials were not designed for a study of intensity, they nonetheless provide strong indication that it is not correlated reliably with stress.

3.4.1 Data and analysis

The current section analyses the original, unedited data elicited by Production Experiment I, i.e. the same 947 utterances used in Chapter 2. The lexical materials are repeated in Table 3.12, otherwise see section 2.2 for documentation. Measurements of minimum, maximum and mean intensity were conducted for all syllable nuclei in decibels (dB). Measurements from nuclei realised partially or completely with non-modal voice quality were excluded from all analyses, since these segments consistently showed lower intensity values, a tendency found in several languages according to Gordon and Ladefoged (2001, p. 397)—on the distribution non-modal voice quality in the current data and its function in marking information structure, see Chapter 2. The measurements of mean intensity were evaluated for the complete data set. Additionally, a subset of the measurements were evaluated to assess the effect of stress directly. The nuclei chosen for this targeted comparison are underlined in Table 3.12. As indicated by the font colours, the subset materials consisted of five instantiations of identical nuclei—[a, aː, oi, ø, u]—of which one was stressed and the other one was unstressed. Thereby, the unstressed member of each nucleus pair was always part of a syllable whose position directly preceded the position of the syllable containing the stressed member of the pair.

The reason for choosing this subset for the investigation of stress was that within the data set as a whole, the factors stress and position were conflated. Finnish primary stress is always located on the first syllable and all words in the present material were disyllabic. Thus, primary stressed syllables were systematically located earlier in the sentence and the materials contained no stressed and unstressed versions of an identical syllable nucleus in the same position. For example, a more ideal set of materials for the investigation of

Table 3.12: Lexical material used in Production Experiment I (=Table 2.1) and pairs of syllable nuclei used for targeted comparison (underlined).

	Subject	Verb	Object	Translation
1	Niilo	maalaa	liinaa	"Niilo (m) paints (a) cloth."
2	Moon$\underline{\text{a}}$	liim$\underline{\text{aa}}$	n$\underline{\text{aa}}$vaa	"Moona (f) glues lichen."
3	Maini	neul$\underline{\text{oi}}$	l$\underline{\text{oi}}$mea	"Maini (f) knitted a blanket."
4	Väin$\underline{\text{ö}}$	jauhoi	lyijyä	"Väinö (m) ground lead."
5	Jani	t$\underline{\text{ö}}$ni	lavaa	"Jani (m) pushed a platform."
6	Jimi	l$\underline{\text{u}}$ki	menua	"Jimi (m) read the menu."
7	Man$\underline{\text{u}}$	h$\underline{\text{a}}$li	lelua	"Manu (m) hugged a toy."
8	Jali	nyki	lanaa	"Jali (m) tugged a levelling drag."

stress might contain a pair of identical syllables as those underlined in (19a), where boldface marks primary stress.[8] However, since this section presents a post-hoc study, the materials were not designed for the investigation of stress. Instead of comparing stressed and unstressed syllables in identical position, this section therefore compares stressed and unstressed versions of the same syllable nucleus in adjacent positions, as schematically illustrated in (19b). For each nucleus, the unstressed version was earlier than the stressed version, so that the analysis tested whether a positive effect of stress (higher intensity on stressed syllables) would outweigh a negative effect of position (lower intensity later in the sentence). Thus, for [a, ø, u] the unstressed member of the pair was the nucleus of the second syllable of the first word in the sentence, while the stressed member was the nucleus of the first syllable of the second word. For [aː, oi], the unstressed member of the pair was the nucleus of the second syllable of the second word (i.e. the overall fourth syllable in the sentence), whereas the stressed nucleus was in the first syllable of the second word (i.e. the fifth syllable in the sentence).

(19) a. $\boldsymbol{\sigma}\sigma\ \boldsymbol{\sigma}\sigma\underline{\sigma}\ \boldsymbol{\sigma}\sigma$
$\boldsymbol{\sigma}\sigma\ \boldsymbol{\sigma}\sigma\ \underline{\boldsymbol{\sigma}}\ \boldsymbol{\sigma}\sigma$

b. $\boldsymbol{\sigma}\underline{\sigma}\ \boldsymbol{\sigma}\sigma\ \boldsymbol{\sigma}\sigma$
$\boldsymbol{\sigma}\sigma\ \underline{\boldsymbol{\sigma}}\sigma\ \boldsymbol{\sigma}\sigma$

Statistical analyses were performed by fitting linear mixed-effects models as implemented in R, with the maximum number of iterations set to 1000 (R Development Core Team, 2010; Baayen, 2008). Models were compared with the ANOVA function. In interpreting the best-fitting converging model, effects with $t > 2$ were treated as significant.

[8]This example is intended to convey the general idea of stressed vs. unstressed realisations in the same position. A well-designed study would need to consider foot structure carefully to distinguish between primary stress, secondary stress and unstressed syllables.

3.4.2 Results

The best converging model of mean intensity in the complete data set showed several interactions and was overall relatively complicated (see Table 3.13). Most importantly, the model included a significant effect of stress: Stressed syllables had an overall higher mean intensity than unstressed ones. This effect interacted with focus condition and was significantly stronger for narrow focus words and significantly or marginally weaker for unfocused ones. Focus condition also had a main effect in that syllable nuclei in narrowly focused words had an overall higher mean intensity. A further main effect was that of position, showing that intensity was lower for vowel nuclei of verbs and objects than for those of the sentence-initial subjects. Again, there were interactions with focus condition, indicating that the differences between narrow focus, broad focus and unfocused condition were larger later in the sentence.

In contrast, the connection between the factors stress, position and focus condition emerges clearly from Figure 3.15. The effect of position is visible at first glance, with higher mean intensity values in earlier syllables. This effect was however modulated by focus condition. While syllable nuclei of unfocused words exhibited a rather steep and continuous fall of intensity, i.e. a clear negative correlation with position, the effect was less strong in broad focus, where the largest intensity drops are located between nuclei of the first and second and of the final and pre-final syllable, respectively. Also, it seems as if the effect of position were attenuated by stress in broad focus, since mean intensity values of unstressed and following stressed syllables were almost identical (compare S2 to V1 and V2 to O1), although the overall small differences in sentence-medial syllables make it difficult to gauge the importance of this observation. The distinction between stressed and unstressed syllable nuclei was largest in narrow focus. Nuclei of stressed (and thus first) syllables in narrowly focused words reached very similar mean intensity values regardless of their position in the sentence. Also, the sentence-final drop was larger than in other focus conditions.

These findings underline the importance of separating stress and sentence position for allowing an investigation of the effects of stress itself. Since the effect of position on intensity was so prevalent in the data set as a whole, conflating both factors by comparing consistently earlier stressed syllables with consistently later unstressed syllables will most certainly distort the analysis.

Indeed, a rather different picture emerged for the targeted comparison subset. Stressed syllables did not show systematically higher intensity than unstressed ones. This can be seen in Figure 3.16 and is confirmed by the linear mixed-effects models of mean, maximum and minimum intensity in Tables 3.14–3.16. The three models were overall very similar and none of

Table 3.13: Coefficients of the linear mixed-effects model of mean intensity (in dB) in the complete data set, with random effects of stress by syllable nucleus and of stress, focus condition, position and trial by subject. The factor trial was centred to its median. CF indicates contrastive focus, IF information focus on the narrow focus constituent.

	Estimate	Std. Error	t value
(Intercept)	61.5833	0.7600	81.0286
Stressed	1.1304	0.3027	3.7344
Narrow focus (CF)	−2.2025	0.2780	−7.9232
Narrow focus (IF)	−1.7146	0.3179	−5.3932
Unfocused (CF)	−0.2688	0.2500	−1.0754
Unfocused (IF)	−0.2232	0.2394	−0.9323
Position 2	−1.6267	0.3011	−5.4025
Position 3	−2.1230	0.4797	−4.4253
Quantity 2	1.1179	0.3457	3.2338
Trial	−0.0005	0.0031	−0.1595
Stressed : Narrow focus (CF)	3.2319	0.2640	12.2411
Stressed : Narrow focus (IF)	2.0815	0.2574	8.0861
Stressed : Unfocused (CF)	−0.3762	0.2069	−1.8180
Stressed : Unfocused (IF)	−0.4802	0.2047	−2.3461
Narrow focus (CF) : Position 2	1.1582	0.2665	4.3454
Narrow focus (IF) : Position 2	1.3218	0.2613	5.0594
Unfocused (CF) : Position 2	−1.1524	0.2215	−5.2021
Unfocused (IF) : Position 2	−0.8533	0.2204	−3.8720
Narrow focus (CF) : Position 3	0.4290	0.3358	1.2777
Narrow focus (IF) : Position 3	0.6797	0.3341	2.0346
Unfocused (CF) : Position 3	−2.8530	0.3202	−8.9093
Unfocused (IF) : Position 3	−2.1126	0.2967	−7.1196
Narrow focus (CF) : Quantity 2	−0.8685	0.2560	−3.3921
Narrow focus (IF) : Quantity 2	−0.4011	0.2533	−1.5839
Unfocused (CF) : Quantity 2	0.4026	0.2167	1.8580
Unfocused (IF) : Quantity 2	0.2568	0.2121	1.2106
Position 2 : Quantity 2	−0.6608	0.2595	−2.5460
Position 3 : Quantity 2	−1.7968	0.3225	−5.5722

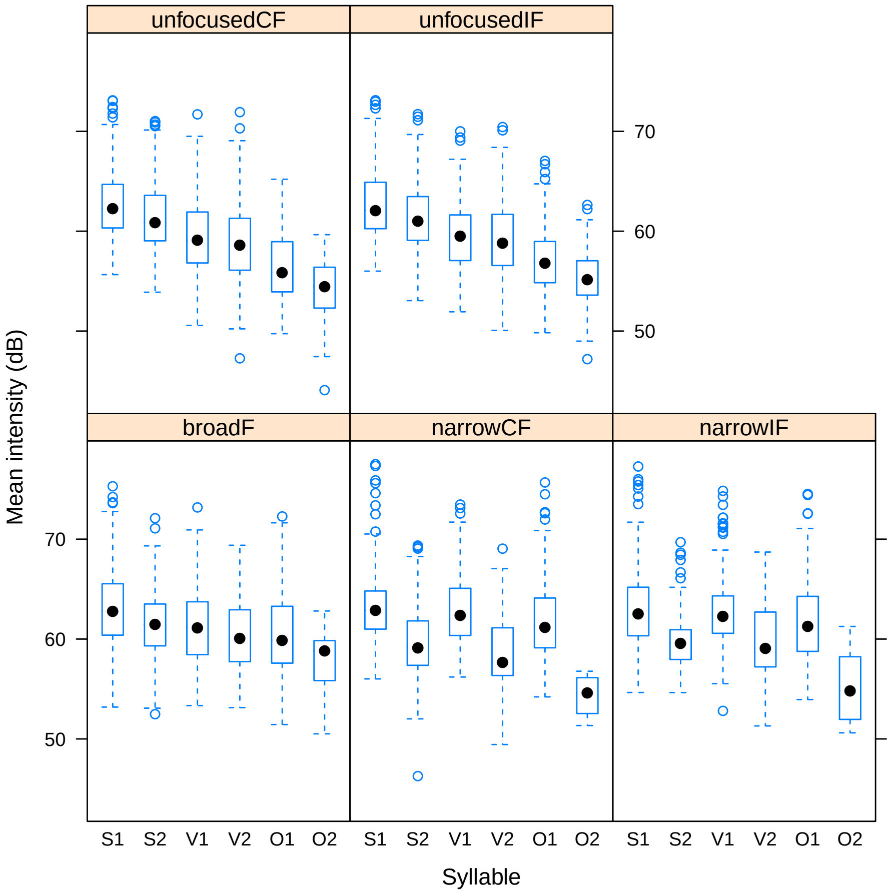

Figure 3.15: Mean intensity of syllable nuclei in different positions by focus condition in the complete data set. “S1” and “S2” indicate the first and second syllable of the subject, respectively, “V1” the first syllable of the verb etc.

them included a significant effect of stress.[9] Instead, the occurrence of high intensity values was restricted to a specific condition: Nuclei of stressed syllables in narrow focus words were marked with high intensity, while intensity was very low in the following unstressed syllables of the same words. Thus, a large difference between stressed and unstressed nuclei can be seen for the narrow focus condition in all panels of Figure 3.16. This was reflected by significant interactions in all three models, combined with a negative main effect of narrow focus. In contrast, nuclei of stressed and unstressed syllables in broad focus did not differ much in their mean and maximum intensity values, as the left panels in Figure 3.16 illustrate. Intensity minima in broad focus, as well as mean values, maxima and minima of syllable nuclei in unfocused words even showed lower values in stressed than in unstressed syllables. As stressed syllables were consistently later than unstressed syllables in this subset of the data, this is probably connected to the effect of position. Syllable nuclei of sentence-medial verbs and sentence-final objects still had significantly lower minimum and mean intensity and marginally lower intensity maxima than syllable nuclei of sentence-intial subjects, as the main effects of position in Tables 3.14–3.16 indicate.

3.4.3 Discussion

The analyses in the preceding section suggest that intensity is not a correlate of stress in Finnish. When the effect of position was controlled for, intensity was not higher in stressed than in unstressed syllables outside narrow focus. At the same time, the current findings confirm and expand Vainio and Järvikivi's (2007) analysis of prosodic focus marking, which likewise found an effect of (narrow) focus on intensity. The present results detail this finding: Words in narrow focus carried a high intensity peak on the first syllable, which was followed by a fall to the second syllable, thus paralleling the pitch fall characteristic of narrow focus (see section 2.4.2).

The absence of an effect of stress on intensity matches the results of Tuomainen et al. (1999), who performed a regression analysis on speech segmentation data including stressed and unstressed syllables. They concluded that the faster reactions to stressed syllables could not be predicted by differences in intensity. The current findings are also in line with Ylitalo (2009)'s conclusion that "is not known to date whether stress in Northern Finnish or in some other variety of Finnish is realised by some means other than segment durations" (Ylitalo, 2009, p. 16).

[9] Note that for this subset, effects of quantity could not be investigated. Models including both the factors position and quantity did not compute, since quantity was entirely correlated with sentence position. All analysed quantity 1 nuclei occurred in position 1 (unstressed) and 2 (stressed), whereas all quantity 2 nuclei were in position 2 (unstressed) and 3 (stressed).

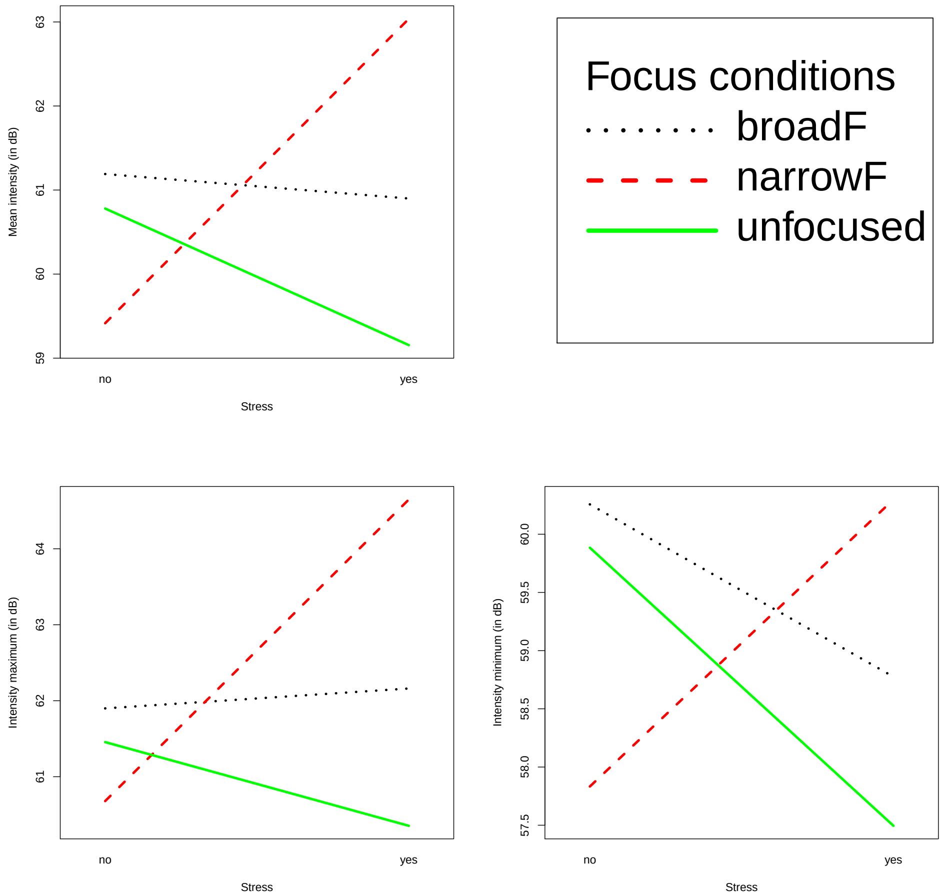

Figure 3.16: Interaction between stress and focus condition for mean intensity (top panel), intensity maxima (bottom left panel) and intensity minima (bottom right) in dB. Values for unstressed syllables are plotted on the left, those for stressed syllables on the right within each interaction plot.

Table 3.14: Coefficients of the linear mixed-effects model of mean intensity (in dB) in the targeted comparison subset, with random effects of stress by vowel and of stress, position and trial by subject. The factor trial was centred to its median. CF = contrastive focus, IF = information focus.

	Estimate	Std. Error	t value
(Intercept)	61.5662	0.7761	79.3240
Stressed	0.6255	0.7695	0.8130
Narrow focus (CF)	−2.1802	0.3718	−5.8643
Narrow focus (IF)	−1.4020	0.3655	−3.8356
Unfocused (CF)	−0.3296	0.2712	−1.2152
Unfocused (IF)	−0.3193	0.2687	−1.1881
Position 2	−1.1978	0.5694	−2.1036
Position 3	−2.5830	1.2629	−2.0454
Trial	0.0000	0.0034	−0.0050
Stressed : Narrow focus (CF)	4.4850	0.5083	8.8234
Stressed : Narrow focus (IF)	3.5749	0.5040	7.0927
Stressed : Unfocused (CF)	−1.1035	0.4241	−2.6017
Stressed : Unfocused (IF)	−1.0084	0.4137	−2.4374

Table 3.15: Coefficients of the linear mixed-effects model of intensity maxima (in dB) in the targeted comparison subset, with random effects of stress by vowel and of stress, position and trial by subject. The factor trial was centred to its median. CF = contrastive focus, IF = information focus.

	Estimate	Std. Error	t value
(Intercept)	62.2118	0.7902	78.7282
Stressed	1.0428	0.6209	1.6796
Narrow focus (CF)	−1.4230	0.3772	−3.7723
Narrow focus (IF)	−1.0333	0.3709	−2.7855
Unfocused (CF)	−0.3261	0.2753	−1.1842
Unfocused (IF)	−0.3344	0.2728	−1.2259
Position 2	−1.0733	0.5874	−1.8272
Position 3	−1.8779	0.9572	−1.9619
Trial	−0.0004	0.0034	−0.1305
Stressed : Narrow focus (CF)	4.0414	0.5162	7.8295
Stressed : Narrow focus (IF)	3.4788	0.5120	6.7940
Stressed : Unfocused (CF)	−1.1214	0.4314	−2.5996
Stressed : Unfocused (IF)	−1.1163	0.4206	−2.6543

Table 3.16: Coefficients of the linear mixed-effects model of intensity minima (in dB) in the targeted comparison subset, with random effects of stress by vowel and of position and trial by subject. The factor trial was centred to its median. CF = contrastive focus, IF = information focus.

	Estimate	Std. Error	t value
(Intercept)	60.7490	0.7626	79.6636
Stressed	−0.0084	1.1859	−0.0071
Narrow focus (CF)	−3.0696	0.4451	−6.8961
Narrow focus (IF)	−1.9524	0.4380	−4.4571
Unfocused (CF)	−0.4074	0.3245	−1.2553
Unfocused (IF)	−0.2285	0.3215	−0.7106
Position 2	−1.5513	0.4439	−3.4949
Position 3	−4.6600	2.0851	−2.2349
Trial	0.0007	0.0035	0.2047
Stressed : Narrow focus (CF)	4.6286	0.6101	7.5865
Stressed : Narrow focus (IF)	3.6123	0.6064	5.9566
Stressed : Unfocused (CF)	−0.6209	0.5098	−1.2180
Stressed : Unfocused (IF)	−0.7758	0.4971	−1.5605

In fact, I would like to argue that so far, no unquestionable phonetic correlates of stress have been established for Finnish. Note that while Ylitalo (2009) confirms the durational marking of primary stress, she finds no duration effects of secondary stress. Segment durations have a crucial role in marking quantity differences, which are not only signalled on the respective segment, but influence durations of neighbouring segments (Wiik and Lehiste, 1968; Lehtonen, 1970; Suomi, 2009). Since most Finnish word stems are disyllabic (see Suomi et al. 2008, p. 69, also see Karlsson 2005), it is arguably especially important to distinguish long and short segments in the first two syllables of words, which is where durational adjustments are in fact localised (Suomi, 2007b).[10] It therefore stands to reason that overall longer durations in the first two syllables serve to enhance the crucial differences between segment quantities in word stems instead of being a marker of primary stress. Alternatively, longer durations could possibly be analysed as a word boundary marker. At any rate, longer word-initial durations do not necessarily have to be interpreted as a correlate of stress.

Phonetic correlates that are often discussed as differentiating stressed from unstressed syllables are f_0, intensity, duration and spectral tilt, corresponding to the psychoacoustic dimensions pitch, loudness, length and qual-

[10]However, Suomi and Ylitalo (2004) describe the durational effect of stress as localised on the first two moras. See Karlsson (2006, 2007) for a critique of this description.

ity (Fry, 1958). While pitch is modelled as accents associated with stressed syllables rather than to be viewed as a correlate of stress itself in the AM-approach, the relative importance of the other parameters has been discussed and differences have been found between the languages studied. For American English, Okobi (2006) found that stress reliably correlated with syllable duration, spectral tilt and noise at high frequencies (for further discussion, also see references therein).

While more complex spectral analyses have not been performed, Suomi et al. (2008) state that stress "has at most very little effect on [perceived, AA] vowel quality" (Suomi et al., 2008, p. 22) in Finnish. Also, the aforementioned regression analysis by Tuomainen et al. (1999) not only failed to establish intensity as a factor predicting the effects of stress differences. Additionally, the authors preformed the same analyses with f_0, duration and spectral tilt as potential predictors. Only f_0 reliably correlated with stress effects, while this was not the case for duration and spectral tilt.

In light of the present section's findings, I would like to tentatively advance the idea that stress has no phonetically correlates in Finnish at all. Obviously, this suggestion will have to be tested by future research. If it were confirmed, however, it would imply that Finnish intonation cannot be described as a classical intonation language with pitch accents associated with stressed syllables. This would constitute strong support for the phrase tone analysis advanced here, although the hypothesis that Finnish is a phrase language does not hinge on the issue of stress (cf. the discussion in Chapter 6).

3.5 Conclusion

This chapter has provided clear evidence against the nuclear accent hypothesis. Instead of showing two different accents, nuclear and non-nuclear ones, the analyses of alignment in the edited data from Production Experiment I indicated that peak timing is highly uniform in the present data set and can be modelled as an H tone realised usually during the first syllable at a fixed distance from the beginning of the nucleus (see section 3.2). This tone was analysed as associated with the p-phrase (H_p). Observed late f_0 maxima seemingly deviating from this pattern were explained as realisations of tones associated with the higher-ranking i-phrase (H_i). The results showed that H_p and H_i usually co-occur and are not in complementary distribution. Apart from the presence or absence of H_i resulting in final rises, the results of the present experiment thus showed the existence of only one tonal contour in spite of the induced variation in quantity, position / grammatical function and focus condition.

In a small pilot perception study presented in section 3.3, manipulations of peak f_0 significantly affected whether participants perceived the respective constituent as narrowly focused, but manipulations of peak timing did not affect the judgements. This is taken as further evidence that the overall earlier peak timing in the unedited / minimally edited data from Production Experiment I was not indicative of a phonologically meaningful difference as proposed by the nuclear accent hypothesis contrasting H*L and LH* accents. Instead, while the pilot study—naturally—did not conclusively prove the uniformity hypothesis, it is congruent with assuming the existence of only one phonological specification H_p for the majority of f_0 maxima discussed here (H_i constituting the second, minority option).

Section 3.2.4 argued that the finding of one uniform tonal contour opens up at least two possible analyses: first, that there is only one type of accent, second, that these peaks are not part of an accent, but realisations of phrase tones. While the two options are difficult to distinguish empirically, section 3.4 brought forward some findings that might tip the scales in favour of the phrase tone analysis. It did not find evidence for a marking of word-initial stress by intensity and thus suggested that Finnish differs from prototypical intonation or stress-accent languages.

Chapter 4

Production Experiment II: Syllable structure and information structure

4.1 Introduction

This chapter presents the results of Production Experiment II, designed to test and extend the hypotheses regarding tonal analysis, quantity and focus marking as derived from Production Experiment I. The analysis of intonation and phrasing concentrates on the p-phrase, which emerged as central in shaping the pitch contours of Finnish utterances. The current chapter confirms the postulation of the phrase tones H_p and L_p. It further establishes the alignment of these tones in more detail, arguing that they are aligned with the beginning and end of the p-phrase, respectively. In doing so, it also re-addresses the question of variation in peak alignment and applies the uniformity hypothesis discussed in Chapter 3 to a different data set.

The primary interest in the design of Production Experiment II was to allow a closer investigation of quantity and—in connection with it—syllable structure in general. Thus, a greater number of initial syllable types was elicited compared to Production Experiment I, as well as a difference in second syllable vowel quantity, which was either long (Q2) or short (Q1). Also, the materials contained trisyllabic in addition to disyllabic words. This variation made it possible to localise intonational and other effects more precisely. It also allowed some tentative observations regarding the mora as a timing unit and the tertiary classification of syllable weight.

Concerning information structure, the design only induced a difference between broad focus and narrow focus on the target words. Nevertheless, prosodic means of focus marking emerged from the data, which again provided a more detailed account of their timing.

In addition to height and timing of f_0 turning points, duration and voice quality were again examined as a prerequisite.

4.2 Methods

4.2.1 Materials

The material was based on five sets of target word pairs contrasting in first syllable structure. Table 4.1 illustrates the first syllable contrast with minimal pairs in the basic nominative case form, which was always disyllabic with the second syllable being CV. A subscript g marks geminate consonants.

Table 4.1: First syllable contrasts embedded in the target words of Production Experiment II.

Contrast set	*Compared syllable structures*	*Example for item pair*
1	CVVC.CV vs. CV.CV	*saarna* 'sermon' vs. *sana* 'word'
2	CVVC.CV vs. CVV.CV	*saarna* 'sermon' vs. *Saana* (fell in Lapland)
3	CVVC.CV vs. CVC.CV	*paarma* 'horsefly' vs. *Parma* (city in Italy)
4	CVC.CV vs. $\mathrm{CVC_g.C_gV}$	*kulma* 'corner' vs. *kumma* 'strange'
5	CVV.CV vs. CVC.CV	*kuuma* 'hot' vs. *kulma* 'corner'

For each of the five contrasts, there were six pairs of lexical items. As visible from the examples given in Table 4.1, some items were used as members of item pairs in more than one contrast set. This concerned three words: *saarna* 'sermon' (contrast set 1 and 2), *kulma* 'corner' (set 4 and 5) and *vaarna* 'tent peg' (set 1 and 3). As is also visible, the target words were either nouns (45 cases), proper nouns (12 cases) or adjectives (3 cases), since the most important criterion was to find true minimal pairs. All items were real words, although some of the target words are quite infrequent and might have been unknown to some or most of the participants. Due to the near-phonemicy of the Finnish orthography, no difficulty in reading these words should be expected (see Suomi, 2009, p. 401).

All items occurred in three different cases—nominative, partitive and essive—additionally varying the overall syllable structure of the word. Thereby, words in nominative case were always bisyllabic with a short open second syllable (CV) that the partitive marker -*ä* / *a* lengthened to CVV, whereas all essive forms were trisyllabic with the syllable *-nä / na* added to the nominative form. For example, the three forms of the word *sana* 'word' were *sana, sanaa, sanana*, while *kiilto* 'shine' appeared as *kiilto, kiiltoa, kiiltona*.[1]

The target words were embedded in sentences that were as similar as possible for both members of each item pair, both within and across the case / syllable structure conditions. Minimally, the material following the target word and the last segment before the onset of the target word were identical; for six item pairs, the complete pre-target word was the same in all sentences. In particular, it was taken care that morphemes inducing initial doubling for the onset of the target word, like the allative marker *-lle*, appeared in all or none of the pre-target words for a given item pair (on initial doubling, see for example Karlsson and Lehtonen, 1977). All target words appeared in a pre-final position in the sentence, only followed by a short adverbial expression like *pari vuotta* '(for) a few years' or *näkyvästi* 'visibly'. For eight item pairs, target words were part of a larger NP, usually in final position (e.g. *papin mahtava <u>saarna</u>* 'the priest's magnificent sermon' vs. *Lapin mahtava <u>Saana</u>* 'the magnificent Saana of Lapland'), but most targets formed an NP of their own. While the syntactic structure thus differed slightly between the item pairs, it was kept constant within each pair, so that such differences would not affect the comparison between the experimental conditions.

A last variation induced in the material was in information structure: All items appeared in both broad focus and narrow corrective focus. There were thus altogether 360 sentences.

5×	6×	2×	3×	2×	= 360
(1st syllable) Contrast sets	Pairs of lexical items	Members of item pairs	Cases / overall syllable structures	Focus conditions	

The complete list of experimental sentences and the corresponding stimulus questions can be found in Appendix H.

4.2.2 Recording procedure

The 360 experimental sentences were equally distributed among three lists, as also indicated in Appendix H. On each list, both members of each item pair occurred twice in different cases. The case / syllable structure conditions and the focus conditions were also balanced across the lists. Lists 1 and 2 were recorded eight times each with different participants, whereas the third

[1] Again, it can be questioned whether vowel sequences like *oa* in *kiiltoa* 'shine' truly form a diphthong and are thus part of the same syllable (see the discussion in Appendix B).

list was recorded nine times. The 120 experimental sentences on each list were interspersed with 32 filler sentences, which were the same for all three lists.

The recordings took place in the sound-proof recording booth of the Faculty of Behavioural Sciences at the University of Helsinki in October 2010, using a high quality microphone and a sampling rate of 44,100 Hz. The participants stood with their mouth ca. 20 cm from the microphone. The material was individually randomized for each participant and presented on a computer screen in written form sentence by sentence. The monitor presenting the sentences (and, at the beginning of the session, the instructions) was visible through a window in the wall between the recording booth and the neighbouring observation room, about 1 m from the participants' position. The participants were instructed to first read each sentence silently and then utter it as an answer to a question auditorily presented through a loudspeaker in the recording booth. The programme Praat (Boersma and Weenink, 2010) was used for stimulus presentation. To avoid loss of data,[2] I controlled the process from the adjacent observation room, but the speakers could ask for a repetition of the prompt or correct their production. Because each list contained four forms from each item pair and thus very similar words and sentences, slips of the tongue occurred rather frequently for some participants. When speakers did not initiate a correction themselves, I asked them to repeat the respective question-answer pair. Participants were free to take a break any time during the recording session and most sessions included one pause.

For the purposes described above, there was verbal communication between me (the experimenter) and the participants during the recordings through a second loudspeaker in the recording booth, but there was no eye contact and I was usually hidden from the participants' view behind the monitor.

The prompt questions were not recorded, but synthesised by Antti Suni with the GlottHMM speech synthesis system developed at Helsinki University and Aalto University (see Raitio et al., 2011).[3] For the questions inducing corrective focus, a diacritic was added to the word contrasting with the target, causing this word to receive a prominent accent.

[2]In the recordings for Production Experiment I, participants proceeded from one sentence to the next by pressing a button themselves, occasionally skipping a sentence by accident.

[3]Project members: Paavo Alku (principle investigator), Martti Vainio (principle investigator), Antti Suni and Tuomo Raitio; for documentation see the GlottHMM synthesis project homepage (`http://www.helsinki.fi/speechsciences/synthesis/glott.html`).

4.2.3 Participants

Recordings were conducted with 25 native speakers of Finnish (22 female, 3 male). They were all undergraduate or graduate students at the University of Helsinki, between 19 and 41 years old (mean age 24.36 years) and have almost all grown up in the Helsinki metropolitan area. The group was chosen to be as homogeneous and similar to the participants of Production Experiment I as possible with respect to the relevant socio-linguistic characteristics (see Appendix I for details).

4.2.4 Editing and analysis

Editing and analysis of the materials presented here was overall similar to the procedures applied to the data of the first production experiment (see section 2.2.3). Again, a Praat script automatically identified the f_0 maximum (H) and the minima left (L1) and right of it (L2), if present, for the target items in all recorded sentences. I manually verified these detections and—for a subset of the data—annotated relevant f_0 turning points outside the target words (see section 4.4.3 below). At the same time, I manually segmented the target items on the basis of displays of waveform, spectrogram and auditory criteria. Additionally, I marked stretches of speech with non-modal voice quality. As in Chapter 2, the evaluation did not distinguish different types of non-modal voice quality like creaky and breathy voice and only considered the binary distinction between modal and non-modal. For statistical analyses, measured f_0 values were converted to semitones (st) relative to a value of 50 Hz. Unless explicitly stated otherwise, all durational values are in milliseconds (ms).

Altogether, this chapter analyses 2640 items. Because of an error in the script presenting the stimuli, the last 32 items were cut off for the first eleven participants. This resulted in the loss of altogether 293 target items, i.e. almost 10% of the 3000 utterances intended to be recorded (the balance of targets vs. fillers within the last 32 items was different for every participant due to individual randomisation). One further sentence was accidentially skipped in the middle of a recording session. Additionally, 66 sentences were sorted out due to hesitations or slips of the tongue (about 2% of the originally intended number of items).

For the analysis of pitch and alignment, only 2602 sentences were considered. No pitch or only an extremely short stretch was measurable for 10 items, and an additional 28 (about 1% of the data) with very small pitch movements of less than 0.8 st were discarded in accordance with the editing principles applied to the data from the first experiment (see Appendix C).[4]

[4]The process eliminated items with a peak pitch differing from both the preceeding and the following low point (if measurable) by less than 0.8st. As in the previous chapter, also points outside the word itself were taken into account, in cases where I had identified

As in previous chapters, statistical analyses were performed with linear mixed-effects models (R Development Core Team, 2010; Baayen, 2008), interpreting effects with absolute t-values larger than 2 as significant (see section 2.2.4 for more details).

4.3 Hypotheses

The materials were evaluated in terms of target word duration, occurrence of pauses, voice quality, frequency of pitch movements exceeding the target words and alignment and pitch of f_0 turning points. On the basis of the previous chapters and the relevant literature, this section formulates hypotheses for the results, first devoting a separate subsection to alignment before dealing with all other variables. For extrapolating from the results of Production Experiment I, it is crucial that all target words were pre-final within their carrier sentences and only followed by adverbials, which were never narrowly focused and usually not prosodically prominent.[5] They were thus most comparable to the objects in the SVO sentences of Production Experiment I.

4.3.1 Alignment

The preceding chapter confirmed the uniformity hypothesis as stated on page 102. That is, it found that in the edited data from Production Experiment I, the distance of a word's peak (H, analysed as H_p) from the beginning of its first syllable nucleus was not significantly affected by focus condition or first syllable vowel quantity. This chapter hypothesises that the uniformity hypothesis also applies to the current data—including not only a difference between two first syllable vowel quantities, but a larger variety of syllable structures—and can thus be extended:

The uniformity hypothesis (extended) Peak alignment relative to the beginning of the first syllable nucleus is uniform across focus conditions and syllable structures.

The current chapter will test this hypothesis. It will also re-examine which measures of alignment are preferable for the respective tones. As

them during segmentation. For example, when the pitch rise to the target word's peak started on the preceding word the complete rise was taken into account, not only the portion realised on the target word itself. The effect of this decision, however, was minor, as only considering f_0 turning points within the target word would equally have resulted in discarding about 1% of the data (33 items).

Unlike in Chapter 3 (see especially section 3.2.1), no steps were taken to eliminate targets with 'fall from' and 'rise to' contours. During the manual editing, I did not detect a seizable number of these contours on the target words.

[5] This qualification is not based on systematic native speaker judgement, but on my impressions when editing the recordings.

discussed in section 3.2.2, this thesis follows Schepman et al.'s (2006) recommendation to analyse tonal alignment in terms of the absolute difference of a tone from a nearby acoustic landmark. As points of reference for peak alignment, Chapter 3 already discussed the beginning and end of the first syllable nucleus—also Wiik (1988) measured peak alignment relative to the beginning of the word's first vowel. Additionally, peak alignment is measured here relative to two further points, the word onset and the end of the first mora. The beginning of the word is implicit as a point of reference in Suomi's graphs illustrating uniformity (e.g. Suomi, 2007b, 2009), but he and his colleagues state more precisely that the H tone (or end of the rise) is invariably anchored to the end of the word's first mora (Suomi et al., 2003; Suomi, 2005, 2007a,b, 2009). Note that while the other landmarks correspond to segment boundaries that were relatively easy to identify in the acoustic signal, it was not straightforward how to determine the end of the first mora as a measurement point for words beginning with CVV and CVVC syllables, since Finnish long vowels are associated with two moras. Although this does of course not invalidate the reference to abstract phonological units like the mora, it can be difficult or even impossible to establish a correspondence between them and phonetic measures. Nevertheless, Suomi describes the end of the fall not only as phonologically associated, but as aligned with the end of the first mora. For the sake of assessing this description, the measure was therefore included, taking the exact mid-point of the long monophthong vowels as a reference point.

As for the timing of the minima, L1 is expected to be detected close to the beginning of the word in the overwhelming majority of the data, extrapolating from the findings in section 2.4.4 and previous studies (e.g. Suomi, 2007a) and in line with the analysis as due to L_p of the the preceding p-phrase. Section 2.4.4 mostly detected L2 during second syllable vowel nuclei. L2 measurements were thus analysed as realisations of L_p tones associated with the end of the p-phrase. Based on this, L2 should be measurable close to the end of the target words. The current data, including more systematic and extensive variation in syllable structure than that from Production Experiment I, is however more suitable for the investigation of L2 alignment. Different points of alignment have been observed for the end of pitch falls in the literature; while Suomi et al. (2003) and Suomi (2005, 2007a,b, 2009) found L tones anchored to the end of the second mora, Suomi et al. (2008) describe less clear timing around the middle of the third syllable. When comparing these measures, the end of the second mora was also measured in the middle of long vowels when necessary.

In sum, the following hypotheses will be tested:

(20) Hypotheses for alignment

a. The target word's f_0 maximum (H) is uniformly aligned to the beginning of the first syllable nucleus.

b. The f_0 minimum following the maximum (L1) is uniformly aligned to the beginning of the word.

c. The f_0 minimum preceding the maximum (L2) is uniformly aligned to the end of the word.

4.3.2 Duration, pauses, voice quality, pitch movements exceeding the target words and pitch scaling

For duration, the effects of information structure and quantity observed in Chapter 2 are expected to reproduce in the present data set. The effect of quantity is predicted to extend to other positions in the word and generally, additional phonemes can be presumed to cause longer overall duration. The hypotheses regarding duration are thus:

(21) Hypotheses for duration

a. Word duration is longer in narrow focus than in broad focus.

b. Word duration is longer for trisyllabic words than for disyllabic words.

c. Word duration is longer when the second syllable contains a Q2 nucleus than when it contains a Q1 nucleus.

d. Word duration is longer for words with more segments in the first syllable and longer when the first syllable contains a Q2 nucleus than when it contains a Q1 nucleus.

⇒ Predicted durational ranking: CV < CVC, CVC_g, CVV < CVVC

For the occurrence of pauses, the following is predicted based on the findings of Chapter 2:

(22) Hypothesis for pauses

More pauses will appear after narrow focus targets compared to broad focus.

Also for voice quality, the effects of information structure and quantity observed in Chapter 2 are predicted to appear for the current data. On first syllables, a higher percentage of non-modal realisations was found for Q2 than for Q1 vowels in section 2.4.5 (CVV vs. CV syllables). Since this finding was explained in terms of duration, it should extend to other initial syllable types in the present materials. Percentages of non-modal realisations are hypothesised to be high from the second syllable on, with an additional difference between the focus conditions (see the findings for sentence-final objects in Figure 2.17).

(23) Hypotheses for voice quality

a. First syllable: Non-modal realisations are overall low, but more frequent for all other syllable types than for CV syllables.

b. Second and third syllable: Non-modal realisations are more frequent in narrow focus than in broad focus, though frequent in both conditions.

Pitch movements exceeding a single word frequently occurred in the sentences of the first experiment. These materials served to put forth the assumption of H_pL_p tones marking the p-phrase. This analysis predicts that pitch rises to the target words' peaks start before the word onset, even when these words constituted p-phrases of their own. In the current approach, these rises are explained as realisations of the preceding words' L_p and the target words' H_p, while the following falls are analysed as realisations of the H_pL_p tones associated with the target words themselves. Taking into account the prognosis of frequent non-modal endings of target words, rises exceeding the target word should be more frequent than measurable pitch falls exceeding the target words.

(24) Hypotheses for pitch movements exceeding the target word

a. Rises to the peak of the target word frequently start on the preceding word.

b. Falls from the peak of the target word rarely continue into the following word.

Pitch scaling is expected to show the same effects of information structure as in Chapter 2.

(25) Hypotheses for pitch scaling

a. The target word's f_0 maximum (H) is higher in narrow focus than in broad focus.

b. The f_0 minimum following the maximum (L1) is lower in narrow focus than in broad focus.

c. The f_0 minimum preceding the maximum (L2) is higher in narrow focus than in broad focus.

4.4 Results

Results are presented for duration, occurrence of pauses, voice quality and frequency of pitch movements exceeding the target words before turning to alignment and height of f_0 turning points. The section on alignment first presents a general overview and is then in turn sub-divided into separate accounts of peak alignment and alignment of the preceding and following minima, respectively.

4.4.1 Duration

This section analyses the duration of the target words, which was influenced by all experimental factors (see the linear mixed-effects model in Table 4.2; for an illustration of word and segment durations, see Figure 4.4 below).

Table 4.2: Coefficients of the best linear mixed-effects model of word duration (in ms), with random by-item effects of number of syllables, focus condition, second syllable vowel quantity, list, and gender and with random by-subject effects of focus condition, second syllable vowel quantity and trial. N = 2640.

	Estimate	Std. Error	t value
(Intercept)	146.9475	19.0650	7.7077
1st syllable CVC	48.5409	16.7370	2.9002
1st syllable CVC_g	67.3942	20.3290	3.3152
1st syllable CVV	53.1502	17.6369	3.0136
1st syllable CVVC	44.8055	16.8802	2.6543
Narrow focus	21.3772	6.5627	3.2574
Trisyllabic	86.3616	3.8154	22.6350
2nd syllable Q2	81.5474	3.4863	23.3910
Sentence duration	0.0682	0.0034	20.0941
Post-pausal	−14.1732	6.6597	−2.1282
Pre-pausal	22.7934	3.2206	7.0774
List 2	−9.6076	10.1100	−0.9503
List 3	−1.0404	10.1585	−0.1024
Trial	−0.0729	0.0278	−2.6229
Male gender	0.7533	12.6844	0.0594
1st syllable CVC : narrow focus	0.6042	6.2857	0.0961
1st syllable CVC_g : narrow focus	−9.3529	7.6399	−1.2242
1st syllable CVV : narrow focus	4.4862	6.6385	0.6758
1st syllable CVVC : narrow focus	17.7059	6.2987	2.8110

First syllable structure had a significant effect on duration, with first syllables containing more or phonologically longer segments resulting in longer word durations. Table 4.2 shows that all other initial syllable structures lead to longer word durations than CV, for which the overall mean duration was 393 ms. Among words beginning with CVC, CVC_g and CVV syllables, the overall means were very similar, namely 453 ms, 446 ms and 450 ms, respec-

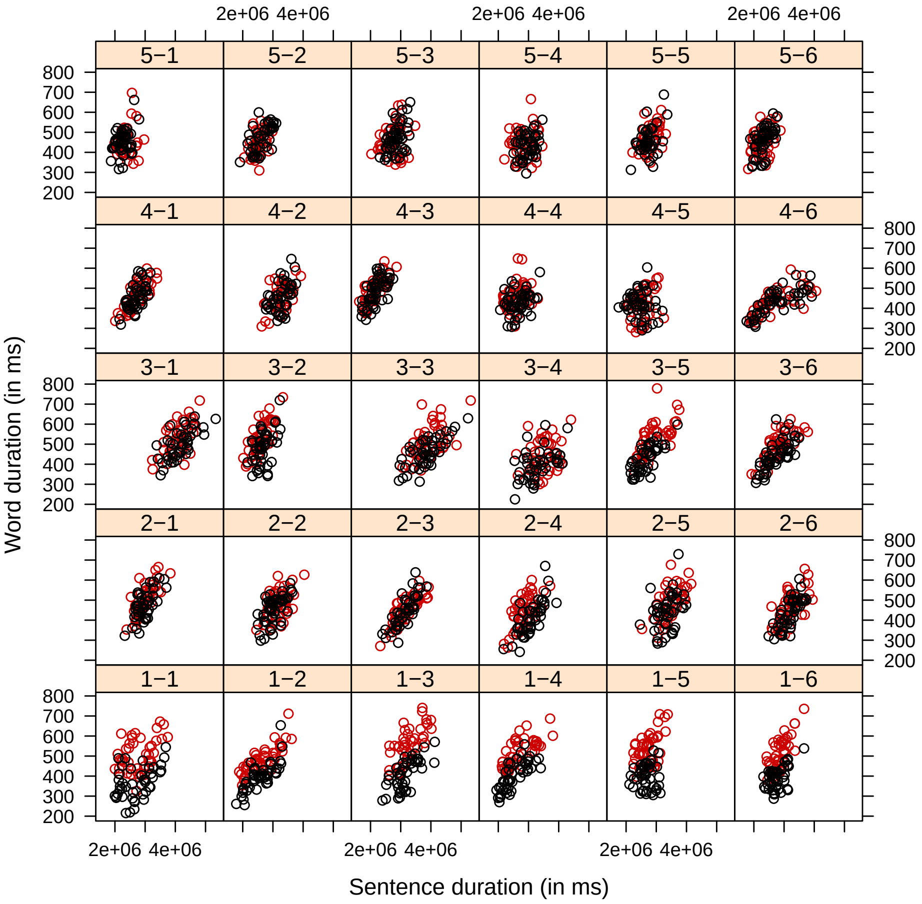

Figure 4.1: Word duration by sentence duration (in ms) for item pairs. Each row of panels shows values for one contrast set, from bottom to top: set 1 (CVVC.X vs. CV.X), set 2 (CVVC.X vs. CVV.X), set 3 (CVVC.X vs. CVC.X), set 4 (CVC.X vs. $\mathrm{CVC_g}$.X) and set 5 (CVV.X vs. CVC.X); see also Appendix H. Within each panel, values for the first member of the respective item pair (e.g. *saarna* 'sermon' in the bottom left panel) are plotted in red, those for the second member (*sana* 'word' in the bottom left panel) in black. The graph plots values from all focus conditions and grammatical cases together.

tively. The largest mean duration, 499 ms, emerged for CVVC.X words, i.e. those with initial CVVC syllables.

Additionally, words in narrow focus had significantly longer durations than those in broad focus. This effect was especially large for words with an initial CVVC syllable, as the interaction in Table 4.2 suggests. Likewise, the difference between CV.X and CVVC.X words and thus the effect of initial syllable structure was larger in narrow focus.

Also, word duration was significantly longer in sentences with a longer overall duration. Since sentence duration was not manipulated in the current materials, the effect cannot be systematically analysed here. However, sentence frames were kept relatively constant within each item pair (see section 4.2.1 and Appendix H). Thus Figure 4.1, which plots values for each item pair in a separate panel, likely illustrates that there was indeed a positive correlation between word duration and sentence duration. The data thus show the opposite of isochrony on the sentence level. A possible explanation is that longer word and sentence durations were both caused by a slower overall speech tempo due to independent factors like structural complexity.

Lastly, the model in Table 4.2 displays an effect of trial, i.e. word durations being shorter later in the experimental session. It is possible that the subjects' overall speech tempo increased as an effect of familiarisation.

4.4.2 Pauses and voice quality

Pauses occurred before 31 and after 193 of the target words, i.e. in 1% and 7% of the analysed sentences, respectively. The difference between focus conditions was very small, with 1% of broad focus and 2% of narrow focus targets being preceded by a pause. Post-target pauses occurred in 7% of broad focus and 8% of narrow focus items. Since the overall number of pauses was so small, no further analysis was attempted here.

Of the 2640 analysed target words, 1455 or 55% were realised partly or completely with non-modal voice quality. Among those, there were only six words that were spoken with creaky / breathy voice throughout. This section thus analyses the distribution of non-modal stretches within target words and the effects of the experimental factors in two ways. First, it investigates the frequency of (partly) non-modal realisations of first, second and third syllables separately. Second, it analyses the extension of non-modal stretches relative to word duration.

For the analysis of voice quality by syllable position, binomial linear mixed-effects models were fit to the data. Positive estimate values indicated more frequent non-modal realisations and negative estimates indicated less frequent non-modal realisations. Models were fit for each syllable separately to keep model complexity manageable (see Tables 4.3–4.5). As Figure 4.2 illustrates with data from essive case targets—the only trisyllabic condition—non-modal realisations were more frequent later in the target words. In the data set on a whole, 19% of first syllables were uttered with non-modal voice quality, compared to 49% of second and 54% of third syllables. Thus, the model for first syllables had the most negative intercept and the model the model for third syllable voice quality had the least negative intercept.

Table 4.3: Coefficients of the best binomial linear mixed-effects model of first syllable voice quality (modal vs. non-modal), with random by-item slopes and random effects of trial by subject. N = 2640.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−4.3454	0.7195	−6.0397	0.0000
1st syllable CVC	2.9075	0.7310	3.9777	0.0001
1st syllable CVC_g	2.1434	0.8595	2.4938	0.0126
1st syllable CVV	2.8912	0.7564	3.8221	0.0001
1st syllable CVVC	3.3806	0.7327	4.6141	0.0000
Trisyllabic	−0.5430	0.1515	−3.5841	0.0003
2nd syllable Q2	−0.4800	0.1508	−3.1835	0.0015
Trial	−0.0003	0.0029	−0.0948	0.9245
Male gender	−3.1737	0.8975	−3.5361	0.0004

Figure 4.2 also shows an effect of focus condition. Narrow focus lead to reliably more non-modal realisations for target words' second and third syllables in essive case and for the data set on a whole, as the models in Table 4.4 and 4.5 confirm. For first syllables, no effect of focus condition was found (see Table 4.3).

Another effect, which was significant for all three syllables, was that of first syllable structure. CV.X words showed less non-modal realisations than words starting with all other syllable types. The difference is most clearly visible for initial syllables in Figure 4.2, but was also significant for second and third syllables (see Tables 4.3–4.5; the only exception was the comparison with CVC_g.X words in Table 4.4). On average across the data set, 3% of initial CV syllables exhibited (partly) non-modal voice quality as opposed to 18% for initial CVC_g syllables, 20% for CVC, 20% for CVV and 24% for CVVC syllables. For the second syllables following them, the respective numbers were 33% vs. 41%, 48%, 55% and 54%. Third syllables were labelled as (partly) non-modal in 42% compared to 57%, 55%, 55% and 56% of the cases when first syllables were CV vs. CVC_g, CVC, CVV and CVVC, respectively.

Table 4.4: Coefficients of the best binomial linear mixed-effects model of second syllable voice quality (modal vs. non-modal), with random by-item slopes and random effects of trial by subject. N = 2640.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−1.3655	0.4563	−2.9926	0.0028
1st syllable CVC	0.8783	0.3098	2.8350	0.0046
1st syllable CVC_g	0.1650	0.3752	0.4396	0.6602
1st syllable CVV	1.4643	0.3295	4.4444	0.0000
1st syllable CVVC	1.2203	0.3120	3.9112	0.0001
Trisyllabic	−0.4592	0.4146	−1.1075	0.2681
Narrow focus	1.0419	0.1317	7.9124	0.0000
2nd syllable Q2	0.9915	0.1334	7.4335	0.0000
Following pause	0.8499	0.2751	3.0892	0.0020
Trial	0.0022	0.0029	0.7551	0.4502
Male gender	−4.3259	1.0630	−4.0694	0.0000
1st syllable CVC : trisyllabic	0.8488	0.4348	1.9521	0.0509
1st syllable CVC_g : trisyllabic	1.2025	0.5251	2.2900	0.0220
1st syllable CVV : trisyllabic	0.5493	0.4582	1.1987	0.2306
1st syllable CVVC : trisyllabic	0.9098	0.4346	2.0934	0.0363
Trisyllabic : narrow focus	−0.4841	0.2216	−2.1841	0.0290
Trisyllabic : pre-pausal	−0.9131	0.4006	−2.2795	0.0226

Table 4.5: Coefficients of the best binomial linear mixed-effects model of third syllable voice quality (modal vs. non-modal), with random by-item slopes and random effects of trial by subject. N = 886.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−0.3998	0.4855	−0.8235	0.4102
1st syllable CVC	0.9293	0.3761	2.4709	0.0135
1st syllable CVC_g	0.9246	0.4629	1.9977	0.0458
1st syllable CVV	0.9615	0.3968	2.4232	0.0154
1st syllable CVVC	0.8710	0.3771	2.3096	0.0209
Narrow focus	0.8745	0.1823	4.7972	0.0000
Trial	0.0049	0.0038	1.2791	0.2009
Male gender	−5.3789	1.3093	−4.1083	0.0000

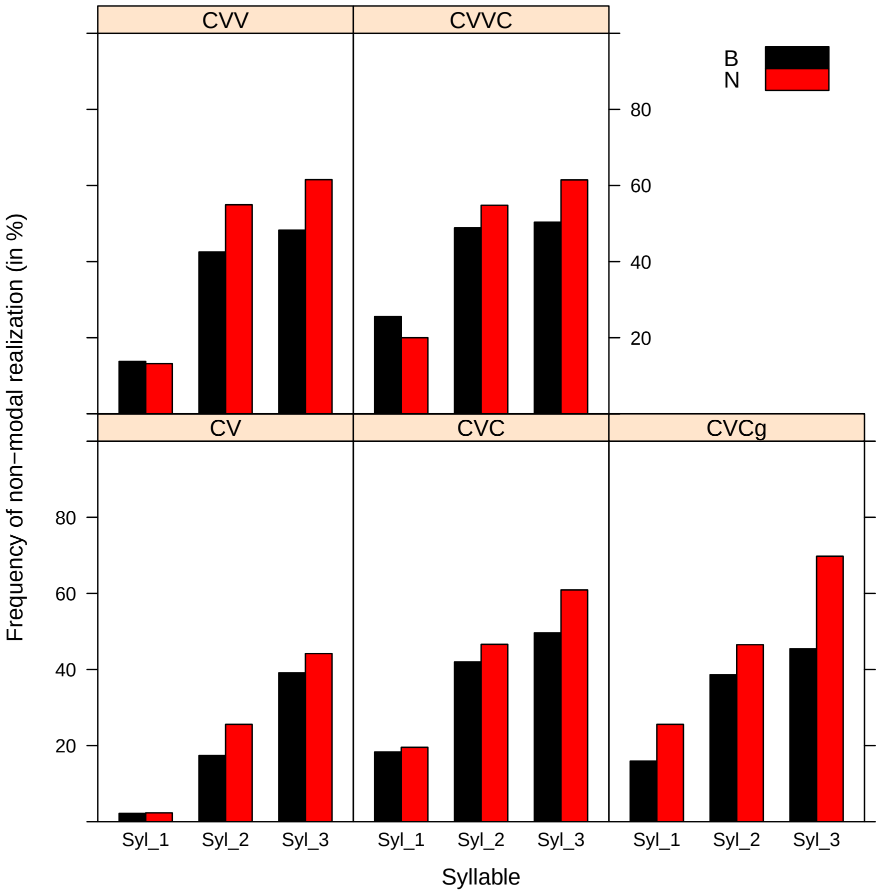

Figure 4.2: Percentage of realisations of first, second and third syllable with non-modal voice quality for target words in essive case (N = 886). Panels distinguish first syllable structures, bars illustrate differences between broad focus (B) and narrow focus (N) realisations, respectively.

Second syllable quantity also had an effect on the voice quality of first and second syllables:[6] Initial syllables were less frequently realised with non-modal voice when they were followed by a second syllable including a Q2 vowel (see Table 4.3), while second syllables were non-modal more often when they contained a long vowel (see Table 4.4).

Additionally, the number of syllables and thus a syllable's position relative to the end of the word also affected the frequency of non-modal voice quality. Word-initial syllables were (partly) non-modal less often in trisyllabic words, i.e. when followed by two other syllables, than in disyllabic ones (see Table 4.3). For second syllables, there was no significant main effect, but interactions with the factors first syllable structure, focus condition and the presence or absence of a pause following the target word. The difference between words starting with CV and those starting with other syllables was larger in trisyllabic than in disyllabic words, since non-modal voice on second syllables was more frequent in CV.CV(V) words than in CV.CV.CV words. Also, the effect of narrow focus leading to more frequent use of non-modal voice was smaller in trisyllabic words, i.e. when the second syllable was not the last syllable of the target word. The third interaction, with the factor called 'pre-pausal' in Table 4.4, evidences that when a second syllable was the last syllable of the word and was followed by a pause, it was realised with non-modal voice more often then when followed by another syllable plus a pause. However, a significant main effect of the factor pre-pausal indicates that second syllables were generally realised with non-modal quality more often when the word was followed by a pause (remember, however, that pauses after the target words were overall quite infrequent in the present data).

Finally, main effects in all three models indicate that the three male speakers produced overall fewer target words with (partly) non-modal voice quality than the female speakers. While men spoke 2%, 6% and 4% of first, second and third syllables, respectively, with non-modal voice quality, women used it for 22%, 56% and 62% of first, second and third syllables, on average.

The second type of analysis focused on the relative duration of non-modal stretches in words which were at least partly realised with non-modal voice. It was calculated here as the percentage of the word duration realised with non-modal voice quality (calculated as *summarised duration of all non-modal stretches during the word* $\times$ 100 / *duration of the word*; henceforth 'percentage of non-modal duration' in short).

The model in Table 4.6 was fit only to the items where this measure was not zero. Its effects evidenced that non-modal stretches were significantly longer in trisyllabic than in disyllabic words relative to word duration. Also,

[6]Recall that trisyllabic words always included CV syllables in second position, so that an effect of second syllable quantity on third syllables could not be investigated with the current materials.

Table 4.6: Coefficients of the best binomial linear mixed-effects model of percentage of word duration realised with voice quality, with random by-item slopes and random effects of focus condition and trial by subject. N = 1455.

	Estimate	Std. Error	t value
(Intercept)	17.9590	2.6420	6.7974
1st syllable CVC	6.2595	2.0515	3.0512
1st syllable CVCg	2.6101	2.4765	1.0540
1st syllable CVV	8.6052	2.1510	4.0006
1st syllable CVVC	7.7437	2.0565	3.7654
Trisyllabic	6.6644	0.8824	7.5525
Narrow focus	0.5845	0.8516	0.6863
2nd syllable Q2	5.7497	0.8727	6.5886
Trial	0.0041	0.0151	0.2732

non-modal voice spanned a larger percentage for words with first and second syllables other than CV.

4.4.3 Pitch movements exceeding the target words

For a subset of 2045 items (77% of the data set), I checked for f_0 movements exceeding the limits of the target words. Of the 2602 items included in the analysis of pitch scaling and alignment, I marked f_0 turning points outside the target word for 78%, i.e. 2020 items.[7] As schematically shown in Figure 4.3, I marked these turning points with the labels 'Lb', 'Hb' and 'La'.

When the rise to the pitch peak of the target word started before the target onset, the beginning of the rise was labelled as 'Lb' (short for 'low before beginning of target word'). In these cases, the initial f_0 minimum of the target word, if present, was identified as L1 (not illustrated in Figure 4.3). A peak preceding the target word, if present, was marked 'Hb' (short for 'high before beginning of target word'), while the f_0 maximum of the target was labelled as 'H'. The label 'La' denoted the end of an f_0 fall continuing after the offset of the target word (short for 'low after end of target word'), with the word-final minimum 'L2' being a part of this fall. Pitch rises only peaking after the target word were not detected in the data, thus the potential label 'Ha' was not employed. Note that phonologically, not all of these tones are

[7] I did not systematically select a subset of the data for this additional level of annotating, but simply made checking for pitch movements exceeding the target a part of the editing process at some point. Because I edited contrast sets 1 and 2 first, very few f_0 turning points outside the targets are marked for these two sets and thus especially for the initial syllable type CV, which only occurred in set 1.

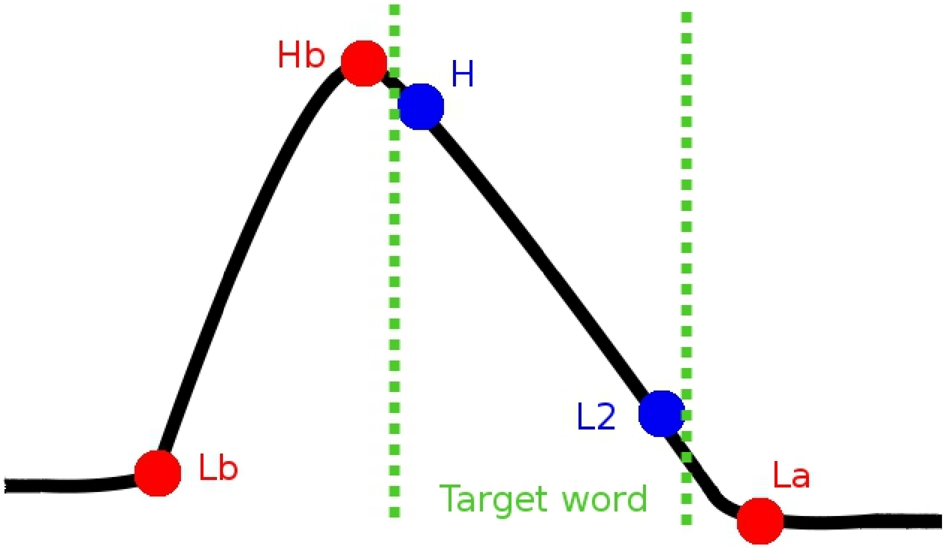

Figure 4.3: Stylised illustration of f_0 turning points outside the target item.

interpreted as associated with the target word. Under the current analysis, Hb and La correspond to H_pL_p tones associated with the p-phrase containing the target word, respectively, while Lb is analysed as an L_p tone associated with the preceding p-phrase (cf. the discussion in section 4.5.1 below).

In looking for turning points outside the target words, I marked only relatively clear cases, where a) the f_0 movement visibly and perceivably exceeded the target word and b) a real turning point could be identified. Thus for example the label 'Lb' was not used when a target starting with a plosive was preceded by lower, but relatively flat pitch, but only when a clear rise could be identified. The following numbers therefore represent a conservative measure.

Within the current data set, Lb, Hb and La never co-occurred for the same target item as illustrated in Figure 4.3.[8] Lb was detected 867 times, i.e. for 42% of the items checked for f_0 movements exceeding the target word and for 33% of all analysed items. For the subset used to analyse pitch scaling and alignment, the respective numbers were 866, 43% and 33%. The marker Hb was used in 21 cases, amounting to 1% of the checked data, as well as to 1% of the data on a whole—the numbers were the same for the pitch analysis subset. For 19 of the 21 items with an Hb point, Lb was also annotated, while Hb and La never co-occured. La did co-incide with Lb in

[8]This was likely due to the fact that target words were always pre-final in broad focus sentences or in narrow focus themselves. Several figures in Chapter 2 show sentence-medial verbs whose pitch is entirely shaped by interpolation between tonal targets associated with neighbouring words.

19 out of the overall 37 times this label was used. Targets with La made up 2% of the data checked for f_0 turning points outside the word and 1% of the complete data set—again, the numbers were the same for the pitch data.

On a whole, f_0 rises starting before the beginning of the target word were much more frequent than falls continuing after the end of the target or a peak occurring before the target onset. Also, for 936 items no word-internal minimum preceding the peak (L1) was identified at all, whereas L2 was not marked for only ten items. As a consequence, statistical analyses in the remainder of this section will consider measurements from Lb points, but not those from Hb or La points, which were rare in the present data. Pitch and alignment of the first minimum are thus always evaluated in two different ways; first, evaluating only word-internal measurements, i.e. L1; second, taking into account Lb measurements when present and substituting these real turning points for the word-internal L1 values. In the latter cases, whether or not I checked for the presence of pitch movements exceeding the target is included as a factor 'Checked for Lb' in the statistical models below.

4.4.4 Alignment

Figure 4.4 depicts average segment durations and mean timing of the word-initial f_0 minimum L1, the maximum H and the final minimum L2 of the target word in all experimental conditions. The panels on the left show values for broad focus, while those on the right illustrate targets in narrow focus condition. The data from the cases nominative (disyllabic, short second syllable vowel: X.CV), essive (trisyllabic, short second and third syllable vowel: X.CV.CV) and partitive (disyllabic, long second syllable vowel: X.CVV) appear in the panels from top to bottom. Within each panel, bars distinguish first syllable types.

For all of these conditions, the mean time of L1 was located during the onset consonant[9] and shows very little variation in Figure 4.4. Evaluating alignment of all individual L1 measurements in the data set, 59% of them were detected on word onset consonants, 40% on first syllable nuclei and less than 1% on first syllable codas or second syllable onsets.

The mean alignment of the peak H was always anchored within the first syllable nucleus. For the data set as a whole, the overwhelming majority of peaks, 89%, were realised during first syllable nuclei. In contrast, only 7% occurred during word-initial consonants, 2% on first syllable codas, 1% on second syllable codas and less than 1% on second and third syllable nuclei each. However, the mean timing per condition as shown in Figure 4.4 exhibited a small, but systematic contrast between initial syllable structures: The mean peak time was later for words starting with CV than for targets

[9] Data from the one pair of items without an onset consonant, *almu* 'alms' and *ammu* 'moo-moo' in nominative (item pair 5 in contrast set 4), was completely excluded from Figure 4.4, but is included in the analyses in section 4.4.4.2.

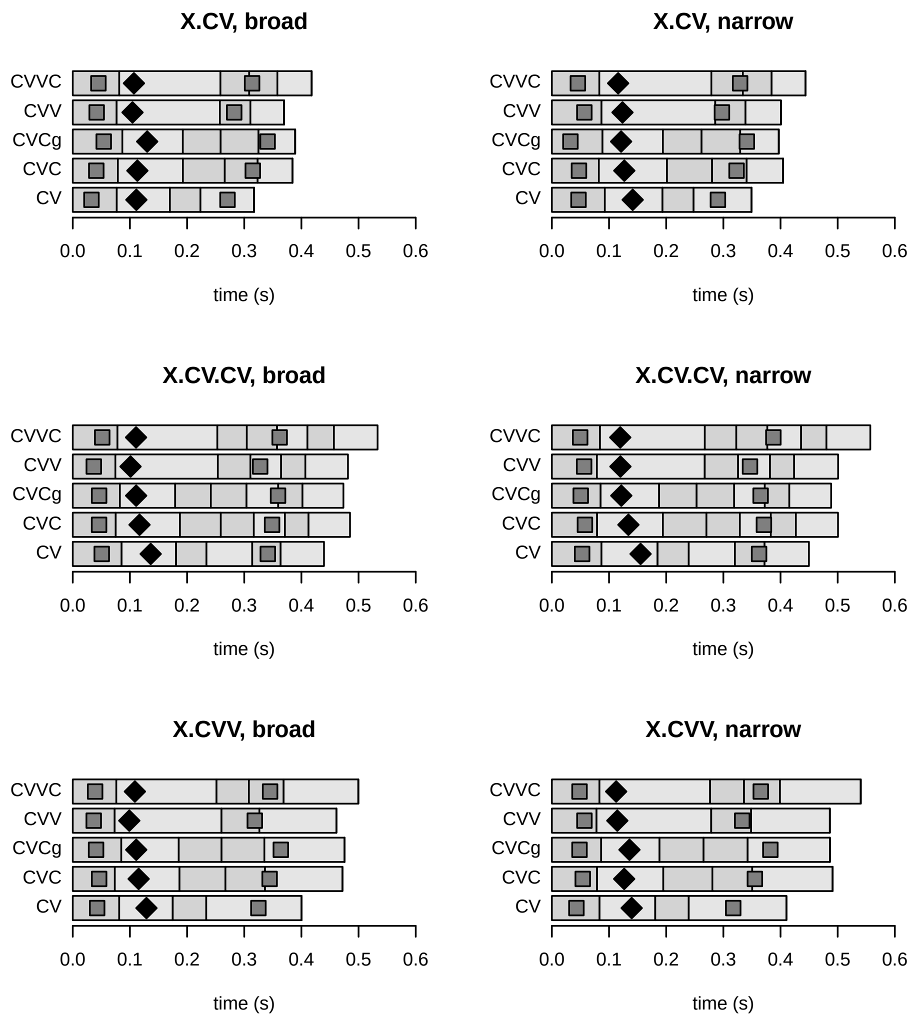

Figure 4.4: Average segment durations (in seconds) and timing of f_0 turning points for all experimental conditions.

with other initial syllable structures in all conditions with the exception of broad focus nominative words. Upon a second look, this contrast was part of a tendency towards later peaks in words with shorter initial syllable nuclei. Mean peaks for words with initial CVVC and CVV were invariably aligned early in the nucleus. Mean peaks times for words with closed initial syllables with short nuclei were equally early in most conditions, but later mean peak times can be seen for others in Figure 4.4. $CVC_{(g)}$-initial words thus seemed to occupy a middle position between words starting with CVV(C) syllables and those starting with CV syllables. Also note that peaks of narrow focus words were generally realised a bit later than those of broad focus words.

The second minima L2 were measured on different segments, most often during second syllable nuclei (38%), while 18% appeared on second syllable onsets, 15% on first syllable codas, 13% on first syllable nuclei, 9% on third syllable nuclei, 6% on third syllable onsets and less than 1% on word-initial onset consonants. For the mean timing of L2, differences between initial syllable structures are visible in the disyllabic conditions (nominative and partitive) in Figure 4.4. For the trisyllabic essive words, the mean times of L2 were almost identical for all initial syllables, but mapped differently to the segmental tier.

While Figure 4.4 concentrates on values measured inside the target words, the alignment of all marked tones is illustrated in Figure 4.5 relative to the beginning of the word. The temporal ranking is Lb $<$ Hb $<$ L1 $<$ H $<$ L2 $<$ La, i.e. Lb was measured earlier than L1 and Hb was earlier than H, while the distribution for La is shifted further away from the beginning of the word than the distribution for L2. Since the label Lb was employed much more often then Hb and La, it is expected that considering f_0 turning points outside the target word has a greater effect for the f_0 minimum preceding the peak (i.e. L1) than for the following minimum or the peak itself. Figure 4.5 confirms this assumption. The distribution of measurements for Lb extends far earlier than that for L1, whereas that of Hb is rather narrow. For L2 and La, the whiskers even overlap (this is of course possible measuring relative to the beginning of the word, since L2 is bounded by word offset).

The following subsections investigate the alignment of each of the three turning points in turn, discussing the best measures of alignment and the impact of the experimental variables. Section 4.4.4.2 also statistically re-examines the effect of including Lb measurements on L1 alignment. Alignment was (as in Chapter 3) always calculated as $T - L$, where T is the time of the tone and L the time of a nearby segmental landmark. Thus, a negative measure indicates that the tone preceded the landmark, while a positive measure indicates that it was realised after the landmark. Since a nearer landmark is to be preferred as a point of reference for leaving less room for variation, it is best to choose a measure that is close to zero for the data on a whole (see Schepman et al., 2006). From a theoretical point of view, a measure $T - L$ being close to zero means that tone T is closely aligned

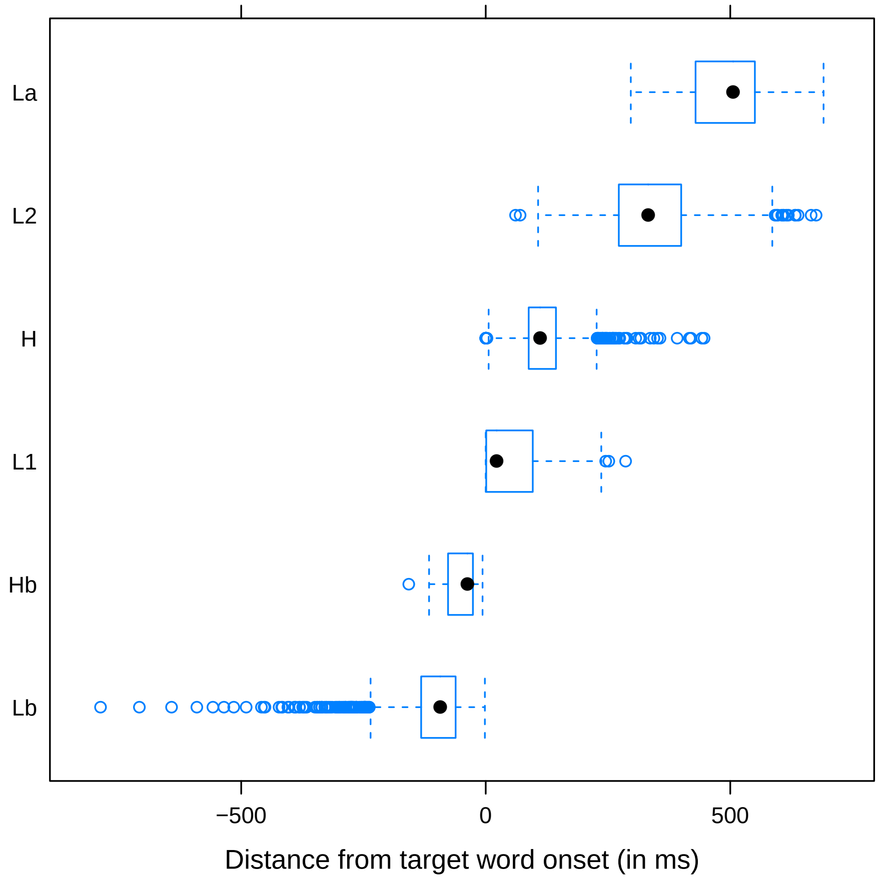

Figure 4.5: Boxplot of timing of for all marked f_0 points relative to the beginning of the target word; distance is given in ms.

with landmark L and thus provides evidence that L is a target of segmental anchoring for T.

4.4.4.1 Peak alignment

Figure 4.6 shows the alignment of the peaks H relative to several potentially relevant landmarks: the beginning and end of the first syllable nucleus, the beginning of the target word and the end of its first mora.

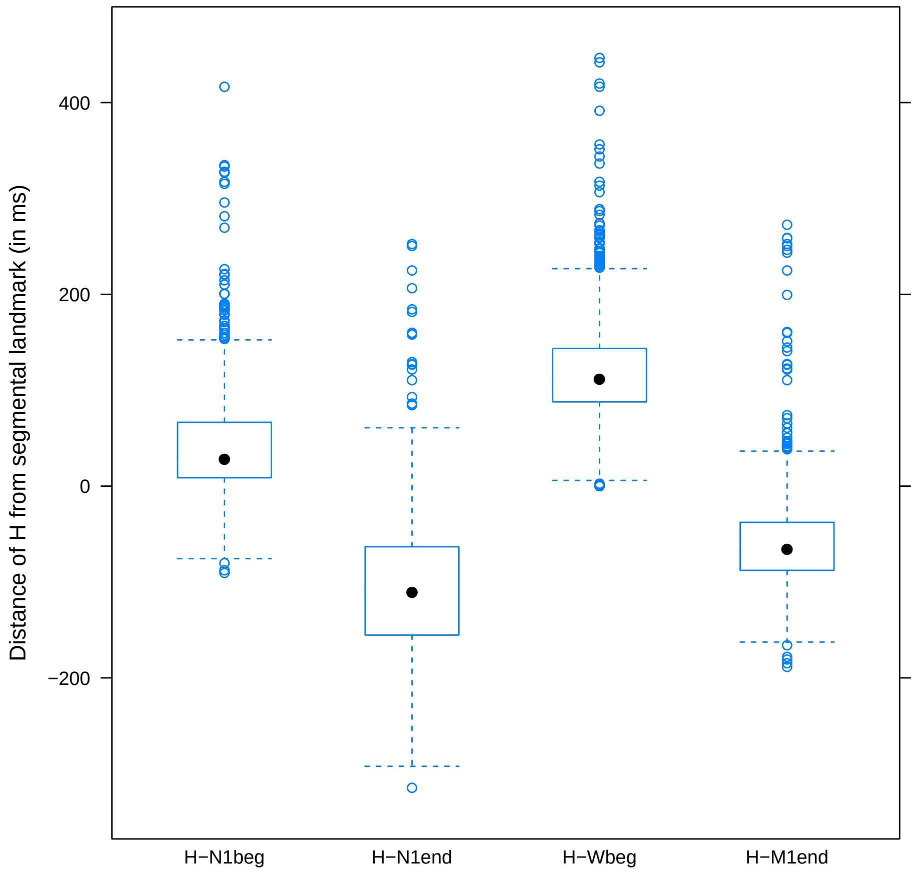

Figure 4.6: Boxplot comparing measures of peak alignment, showing the distance of H from the beginning of the first syllable nucleus (N1beg), the end of the first syllable nucleus (N1end), the beginning of the word (Wbeg) and the end of the first mora (M1end) in ms.

Of the two landmarks compared in Chapter 3, the beginning and the end of the first syllable nucleus, the beginning of the nucleus was concluded to be the better anchor; a finding that Figure 4.6 confirms. The median and lower quartile for $H - N1beg$ were very close to zero (28 ms and 9 ms, respectively), while values for $H - N1end$ were further away from zero. The distance to the beginning of the word was about as large as that to the end of the first nucleus, with the median being 111 ms and −111 ms, respectively. In contrast, the distance to the end of the first mora was smaller (median: −66 ms). Also note that the distribution of $H - N1end$ was wider than that of the other three measures, which is in line with the assumption that the end of the nucleus is not the target of alignment, so that the distance between H and N1end was determined by other factors like nucleus duration. Overall, Figure 4.6 indicates that the measure closest to zero was $H - N1beg$ and that the beginning of the first syllable nucleus is the most likely candidate for a target of segmental anchoring. This can be related to the fact that, as also seen in Figure 4.4, peaks were mostly realised early in the first syllable nucleus.

Peak alignment was therefore evaluated relative to the beginning of the first syllable nucleus, but additional linear mixed-effects models were also fit for distance to the end of the first mora—the second-best anchor point—and to the beginning of the word. These analyses showed effects of all four experimental factors—number of syllables, second syllable quantity, focus condition and first syllable structure—although they were often not significant. Overall, these effects can be summarised as two opposing tendencies: The first tendency was that longer words had later peaks. Peaks were later in trisyllabic than in bisyllabic words and later in narrow focus than in broad focus. There was also a trend to later peak alignments when the second syllable included a Q2 as opposed to a Q1 vowel. The second tendency was already discussed with respect to Figure 4.4 above: targets beginning with a CV($C_{(g)}$) syllable showed later peaks than words with initial CVV(C) syllables.

Table 4.7 gives the mean and standard deviation for the distance from the beginning of the first syllable nucleus in all experimental conditions. Note that the standard deviations from the mean were usually about as large as the mean distances from the point of reference (beginning of the nucleus). Also, differences between the condition means were generally rather small compared to the standard deviations. That said, Table 4.7 illustrates that in all grammatical cases and for both narrow and broad focus, words with initial CV syllables on average always had earlier peaks than words starting with CVV and CVVC syllables, while the relative ranking of CVC($_{g}$) initial words varied. However, the linear mixed-effects model in Table 4.8 indicated that the difference between CV-initial words and those with other initial syllables was not in itself significant for the data set on a whole. It was significant in trisyllabic words (i.e. in essive case), but not necessarily for

Table 4.7: Mean and standard deviation (SD) of the distance of the peak from the beginning of the first syllable nucleus ($H - N1beg$, in ms) in all experimental conditions.

Case	1st syllable	Broad focus		Narrow focus	
		Mean	*SD*	*Mean*	*SD*
Nominative	CV	34	37	49	40
(X.CV)	CVC	37	40	49	43
	CVC_g	60	64	46	48
	CVV	28	38	37	37
	CVVC	26	45	33	30
Essive	CV	51	60	68	51
(X.CV.CV)	CVC	45	45	59	50
	CVC_g	39	39	55	52
	CVV	27	31	41	40
	CVVC	33	53	35	37
Partitive	CV	48	35	56	47
(X.CVV)	CVC	47	53	51	42
	CVC_g	37	42	61	42
	CVV	26	35	36	30
	CVVC	33	43	29	32

the disyllabic intercept, as the interactions suggest.[10] As an illustration, consider Figure 4.4 again, which showed that mean peak times for CV.X words were latest in essive case.

Table 4.8: Coefficients of the best linear mixed-effects model of the distance of the peak from the beginning of the first syllable nucleus (in ms), with random by-item slopes and random effects of focus condition and trial by subject. The factor trial was centred to the median. N = 2602.

	Estimate	Std. Error	t value
(Intercept)	40.1791	7.9561	5.0501
1st syllable CVC	0.6432	8.6983	0.0739
1st syllable CVC_g	4.7314	10.5692	0.4477
1st syllable CVV	−14.0878	9.1646	−1.5372
1st syllable CVVC	−15.4231	8.7602	−1.7606
Trisyllabic	16.4381	4.7879	3.4332
2nd syllable Q2	3.3154	1.7439	1.9011
Narrow focus	8.1685	2.9899	2.7320
Trial	−0.0261	0.0305	−0.8555
1st syllable CVC : trisyllabic	−9.9628	5.4453	−1.8296
1st syllable CVC_g : trisyllabic	−18.4119	6.6812	−2.7558
1st syllable CVV : trisyllabic	−13.0543	5.7795	−2.2587
1st syllable CVVC : trisyllabic	−11.2083	5.4428	−2.0593

Seen from the end of the target word's first mora, words with initial CVC, CVC_g and CVV had significantly earlier peaks than the CV intercept (see Table 4.9). There was no main effect for CVVC syllables, but an interaction showed that the difference between CV.X and CVVC.X words was considerably larger in narrow focus. At the same time, the effect of narrow focus, which overall resulted in significantly later peaks, was smaller for CVVC.X words.

Relative to the beginning of the word, target words with initial CVC, CVV and CVVC syllables showed significantly earlier peak alignment than the CV intercept, while the same effect did not reach significance for CVC_g syllables, as shown in Table 4.10. The differences between initial syllable types were significantly larger in trisyllabic words, as indicated by the interactions at the bottom of the table.

[10] This conclusion was double-checked by fitting separate models for the subsets of trisyllabic and disyllabic items, respectively, as shown in Appendix J.1. A significant effect of first syllable structure was found for trisyllabic words, with CVV.X and CVVC.X words differing from the CV.X intercept. The separate model of disyllabic words only showed a significant main difference for CVC_g.X words, which in turn showed several interactions with other factors.

Table 4.9: Coefficients of the best linear mixed-effects model of the distance of the peak from the end of the first mora (in ms), with random by-item effects of gender and random effects of focus condition and first syllable structure by subject. N = 2602.

	Estimate	Std. Error	t value
(Intercept)	−52.8505	7.1410	−7.4010
1st syllable CVC	−26.6261	7.1964	−3.6999
1st syllable CVC_g	−18.4097	8.8227	−2.0866
1st syllable CVV	−16.7401	7.8179	−2.1413
1st syllable CVVC	−5.6855	7.6487	−0.7433
Trisyllabic	9.1975	5.0435	1.8236
Narrow focus	8.4655	1.7427	4.8577
2nd syllable quantity	4.1236	1.7522	2.3534
Male gender	9.6876	6.9498	1.3939
1st syllable CVC : Narrow focus	−3.9015	5.1898	−0.7518
1st syllable CVC_g : Narrow focus	−6.1171	6.3360	−0.9655
1st syllable CVV : Narrow focus	−5.5921	5.5078	−1.0153
1st syllable CVVC : Narrow focus	−15.4806	5.1878	−2.9841

Table 4.10: Coefficients of the best linear mixed-effects model of the distance of the peak from the beginning of the word (in ms), with random by-item effects of item list and random effects of focus condition and trial by subject. The factor trial was centred to the median. N = 2602.

	Estimate	Std. Error	t value
(Intercept)	132.6868	7.1286	18.6133
1st syllable CVC	−14.8049	5.9926	−2.4705
1st syllable CVC_g	−8.8490	7.3103	−1.2105
1st syllable CVV	−23.6146	6.3516	−3.7179
1st syllable CVVC	−20.2336	6.0180	−3.3622
Trisyllabic	16.0969	5.0221	3.2052
Narrow focus	13.0739	3.7985	3.4419
List 2	−18.3396	7.2774	−2.5201
List 3	−1.6417	7.2116	−0.2276
Trial	−0.0512	0.0347	−1.4754
1st syllable CVC : trisyllabic	−12.3900	5.8044	−2.1346
1st syllable CVC_g : trisyllabic	−21.3061	7.1188	−2.9929
1st syllable CVV : trisyllabic	−16.0210	6.1599	−2.6009
1st syllable CVVC : trisyllabic	−12.2644	5.8034	−2.1133

However, Tables 4.7–4.10 did not distinguish between the contrast sets, but pooled data across the whole data set. Thus, the means and estimates for CV.X and CVC_g.X words were based on the smallest number of separate items, as they each only figured in one of the five subsets (see section 4.2.1). One consequence of this were the exceptionally large mean and standard deviation for broad focus nominative words with an initial CVC_g syllable visible in Table 4.7, which seemed to be caused by two outlier values larger than 200 ms. Therefore, separate models were fit for each subset and for all three measures of peak alignment; see Table 4.11 for a summary of the results and Appendix J.2 for a more detailed discussion. These subset models, which were well balanced in terms of initial syllable contrast, indicated that peaks were significantly earlier in CVVC.X words than in CV.X words relative to the beginning of the first syllable nucleus as well as relative to the end of the first mora. Additionally, the model of the distance of the peak from the beginning of the first syllable nucleus in subset 5 suggested that peaks were earlier in words with initial CVV syllables than in those starting with CVC syllables, but no further effects of initial syllable contrasts emerged from the subset models.

In the data set on whole, words in essive—and sometimes partitive—case generally had later peaks than those in nominative case, as Table 4.7 illustrates. Number of syllables constituted a significant main effect on alignment relative to the beginning of the first syllable nucleus and relative to the beginning of the word (see Tables 4.8 and 4.10, respectively) and a marginal effect relative to the end of the first mora (see Table 4.9). Nevertheless, the subset models indicate that this effect was mainly restricted to sets 1 and 3 (see Table 4.11). The tendency towards later peaks in partitive case, i.e. in words with Q2 vowels in their second syllable, was even more elusive. A marginal effect can be seen for the overall model in Table 4.8 and a significant one in Table 4.9, but the subset models provided very little evidence that second syllable vowel quantity had any weight in a more detailed analysis.

Finally, an effect of focus condition emerges from Table 4.7, which is also visible in Figure 4.4: Peaks were on average realised later in narrow focus than in broad focus in almost all grammatical cases and initial syllable structures.[11] The difference was statistically significant, as Tables 4.8–4.10 demonstrate. Evidence for later peaks in narrow focus words also came from the majority of subset models (see Table 4.11). Additionally, variation in peak timing was slightly larger in broad focus than in narrow focus condition, with a standard deviation of 45 ms vs. 42 ms, 45 ms vs. 39 ms and 47 ms vs. 44 ms for the distance from the beginning of the first syllable nucleus, the end of the first mora and the beginning of the word, respectively. However, one-

[11] One of the two exceptions were the previously discussed $CVC_g.C_gV$ words in broad focus. The second exception, CVVC.CVV words, consisted only of a very small difference between the two focus conditions.

Table 4.11: Summary of linear mixed-effects models of peak alignment in subsets (for more details, see Appendix J.2). An X indicates a significant main effect with $t > |2|$ for the respective factor in the respective model. Interacting factors are marked with an asterisk.

Model		Effect					
Contrast set	Measure	1st syllable contrast	Number of syllables	Focus condition	Second syllable quantity	Trial	Gender
Set 1	H-N1beg	X	X	X			
CV vs. CVVC	H-M1end	X	X				X
	H-W1beg	*	X*	X			
Set 2	H-N1beg						
CVV vs. CVVC	H-M1end						
	H-W1beg			X			
Set 3	H-N1beg		X			X	
CVC vs. CVVC	H-M1end		X		X		
	H-W1beg		X	X		X	
Set 4	H-N1beg			X			
CVC vs. CVC_g	H-M1end						
	H-W1beg			X			
Set 5	H-N1beg	X		X			
CVC vs. CVV	H-M1end		X	X	X		
	H-W1beg			X			

sided paired by-subject and by-item t-tests over SD provided mixed evidence regarding the significance of this difference (for $H-N1beg$: $t(24) = 1.26, p = .1097$ and $t(56) = 1.92, p = .030$; for $H - M1end$: $t(24) = 1.53, p = .069$ and $t(56) = 2.42, p = .009$; for $H - Wbeg$: $t(24) = 2.17, p = .020$ and $t(56) = 1.17, p = .124$, respectively).

4.4.4.2 Alignment of the first f_0 minimum

As mentioned above, rises to the target words' pitch peaks usually started at the beginning of the target word or before. Figure 4.7 compares several measures of alignment for the beginning of the rise, i.e. the first f_0 minimum, as either inside the target word (L1) or before its beginning (Lb). It indicates that of the points of reference compared here, the beginning of the word (Wbeg) was indeed the segmental landmark with which the initial minimum was aligned closest: The median distance for L1, calculated as $L1 - Wbeg$ was 22 ms, while its median distance from the beginning of the first syllable nucleus was −41 ms. Taking into account the frequent f_0 minima preceding the target word, this conclusion was even clearer: The median distance of Lb from the beginning of the word and from the beginning of the first nucleus was −93 ms and −169 ms, respectively. Alignment of the first f_0 minimum was therefore measured relative to the beginning of the target word.

An analysis restricted to the word-internal minimum L1 showed very close and stable alignment to the beginning of the word, as the model in Table 4.12 suggests. It did not evidence an effect of any of the experimental factors (focus condition, first and second syllable structure and number of syllables), but instead infinitesimally later L1 measurements for later trials and very slightly earlier measurements for male participants. For 936 words, i.e. 36% of the evaluated words, no rise to the peak was detectable on the target word itself and L1 was not marked at all.

Table 4.13 shows the best model also considering minima preceding the target word. It also indicated that the experimental factors had little effect, but words with long second syllable vowels showed significantly later first f_0 minima. Minima were also measured significantly later for the three male participants. However, the factor with the largest effect on the measured timing of initial minima was whether or not the editing process marked rises beginning before the target word. That is, when looking for low f_0 turning points before the word, these were often found and occurred at considerable distance from the target word.

4.4.4.3 Alignment of the second f_0 minimum

In contrast to the close to 40% missing values for L1, the second minimum L2 was measured for all but 10 of the 2602 target words analysed in terms of alignment and pitch scaling (less than 0.5% missing values). Figure 4.8

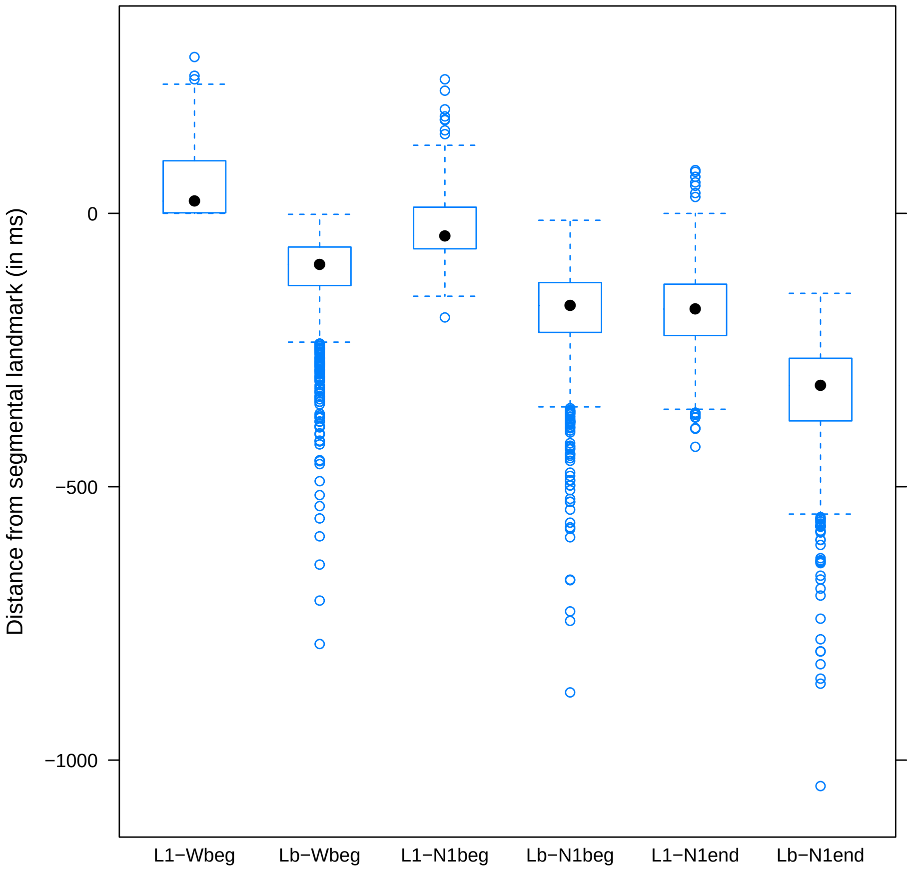

Figure 4.7: Comparing measures of alignment of the first f_0 minimum. The boxplot shows the distance of L1 and Lb, respectively, from the beginning of the word, the beginning of the first syllable nucleus and the end of the first syllable nucleus (from left to right; in ms).

Table 4.12: Coefficients of the best linear mixed-effects model of the distance of L1 from the beginning of the target word (in ms), with random by-item effects of item list and with random by-subject effects of focus condition. N = 1666

	Estimate	Std. Error	t value
(Intercept)	0.0682	0.0062	10.9675
Narrow focus	0.0013	0.0015	0.8570
List 2	−0.0027	0.0025	−1.0718
List 3	0.0011	0.0024	0.4605
Trial	0.0000	0.0000	−2.5403
Male gender	−0.0090	0.0026	−3.4159

Table 4.13: Coefficients of the best linear mixed-effects model of the distance of the beginning of the rise to the peak (Lb or L1) from the beginning of the target word (in ms), with random by-item effects of number of syllables, item list and gender and with random by-subject effects of focus condition. N = 1918.

	Estimate	Std. Error	t value
(Intercept)	28.8040	10.9961	2.6195
Trisyllabic	5.2299	6.8496	0.7635
Narrow focus	1.8519	6.0209	0.3076
2nd syllable Q2	13.0258	5.5119	2.3632
Checked for Lb	−74.9640	9.3025	−8.0584
List2	3.9837	10.3533	0.3848
List3	−0.5464	11.2137	−0.0487
Male gender	40.1388	13.5235	2.9681

compares different measures of alignment for L2, the target word's second minimum. One of them is the distance from the end of the second mora, which Suomi (2007a) identified as the location of the second L tone in his data (see p. 112).[12] For the current data set as a whole, the measure subtracting the end of the second mora from the time of L2 ($L2 - M2end$) was only the second best with a median of 60 ms. The measure closest to zero was $L2 - N2beg$ with a median of 11 ms, so that this section subsequently analyses alignment of the second f_0 minimum relative to the beginning of the second syllable nucleus.

Table 4.14: Coefficients of the best linear mixed-effects model of the distance of L2 (the end of the fall from the peak) from the beginning of the second syllable nucleus (in ms), with random by-item effects of gender and with random by-subject interaction of number of syllables and focus condition and random by-subject effects of second syllable quantity and trial. N = 2592.

	Estimate	Std. Error	t value
(Intercept)	47.4685	6.4186	7.3954
1st syllable CVC	−45.1449	6.1701	−7.3167
1st syllable CVCg	−32.3835	7.5242	−4.3039
1st syllable CVV	−62.8867	6.5366	−9.6206
1st syllable CVVC	−65.5806	6.2050	−10.5689
Trisyllabic	69.8006	4.3143	16.1790
Narrow focus	8.0041	3.9792	2.0115
2nd syllable Q2	42.4276	4.0804	10.3978
% non-modal vq.	−0.9183	0.2139	−4.2923
Trial	−0.0266	0.0382	−0.6968
Male gender	28.2867	8.8900	3.1818
Trisyllabic : narrow focus	18.2861	5.4573	3.3508
Narrow focus : 2nd syllable Q2	8.3840	4.3655	1.9205
1st syllable CVC : % non-modal vq.	−0.8077	0.2097	−3.8521
1st syllable CVCg : % non-modal vq.	−0.6185	0.2552	−2.4235
1st syllable CVV : % non-modal vq.	−0.7414	0.2170	−3.4163
1st syllable CVVC : % non-modal vq.	−1.1912	0.2086	−5.7110
Trisyllabic : % non-modal vq.	−0.6249	0.1272	−4.9119
Narrow focus : % non-modal vq.	−0.5942	0.1063	−5.5894
2nd syllable Q2 : % non-modal vq.	−0.5325	0.1312	−4.0602

[12]For another data set, Suomi et al. (2008, p. 80) found that falls ended around the middle of the third syllable instead. This alternative was not systematically tested here, since L2 was usually measured considerably earlier in the word for the trisyllabic words in the present materials, as Figure 4.4 illustrates.

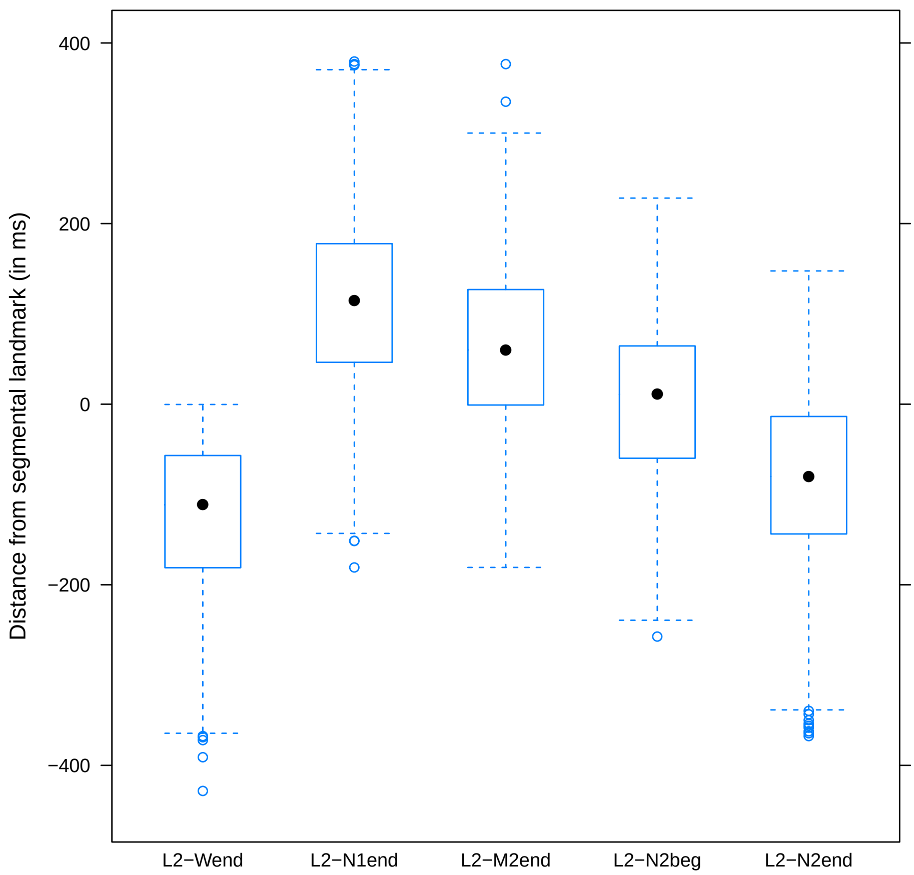

Figure 4.8: Comparing measures of alignment of the second f_0 minimum. The boxplot shows the distance of L2 from the end of the word, the end of the first syllable nucleus, the end of the second mora, the beginning of the second syllable nucleus and the end of the second syllable nucleus (from left to right; in ms).

In terms of this measure, L2 alignment was significantly influenced by all of the experimental variables (see Table 4.14). Compared to the CV.X intercept, words with all other initial syllable types had significantly earlier second f_0 minima. A look at Figure 4.4 confirms that L2 tones in words with longer first syllables were indeed realised overall earlier relative to the beginning of the second syllable nucleus, but that the beginning of second syllable nuclei was also later for words with first syllables containing more segments than for CV.X words. It should be noted that under the current definition of goodness of a measure of alignment, the beginning of the second syllable nucleus was a considerably worse point of reference for CV.X words than for all other initial syllable types. The overall mean distance of L2 from this landmark was 81 ms for words starting with CV syllables, 30 ms when the first syllable was CVC_g, 23 ms when it was CVVC, 11 ms when it was CVC and 9 ms when it was CVV. An additional model of L2 alignment relative to the beginning of the word showed that relative to this less local point of reference, L2 was measured significantly earlier for CV.X words than for words with all other initial syllable types (see Table 4.15; note that this model will not be discussed further here, since the beginning of the second syllable nucleus was determined as the overall best point of reference for L2 alignment above).

Overall, the effects of the experimental variations on L2 alignment relative to the second nucleus onset can be summarised as follows: Second minima were measured later in conditions that induced longer word durations (see section 4.4.1 on word durations). Thus, L2 tones were later in trisyllabic words than for the disyllabic intercept, later in narrow than in broad focus condition and later when the second syllable vowel quantity was Q2 than when it was Q1. These effects were sometimes additionally boosted in combination, as the significant interaction between focus condition and number of syllables and the marginal interaction between focus condition and second syllable vowel quantity in Table 4.14 demonstrate.

Furthermore, voice quality significantly influenced the measured timing of L2. In words with a larger percentage of non-modal realisation, second f_0 minima were measured earlier than in those with a proportionately shorter non-modal stretch.[13] This effect additionally interacted with all experimental factors, reducing effects that caused later L2 measurements and increasing earlier measurements for words with initial syllables other than CV. Note also that L2 alignment was significantly later for male participants, who used non-modal voice quality less frequently overall (see section 4.4.2).

For a further illumination of the influence of voice quality on L2 alignment, consider Figure 4.9. Like Figure 4.4 above, it illustrates mean seg-

[13]Note that in contrast to Table 4.6 analysing the percentage of word duration realised with non-modal voice quality, Table 4.14 contains a model of L2 alignment for the complete data set. Items realised with modal voice throughout were not discarded, instead, the value for percentage of non-modal duration was zero in these cases.

Table 4.15: Coefficients of the best linear mixed-effects model of the distance of L2 (the end of the fall from the peak) from the beginning of the word (in ms), with random by-item interaction of number of syllables and focus condition and random by-subject effects of number of syllables, focus condition, second syllable quantity and trial. N = 2592.

	Estimate	Std. Error	t value
(Intercept)	252.7748	12.5163	20.1956
1st syllable CVC	62.8154	12.3379	5.0912
1st syllable CVCg	65.1933	14.9031	4.3745
1st syllable CVV	46.4152	13.1031	3.5423
1st syllable CVVC	86.2115	12.4267	6.9376
Trisyllabic	94.1255	7.8875	11.9334
Narrow focus	36.7788	4.4544	8.2568
2nd syllable Q2	60.0378	4.1258	14.5519
% non-modal vq.	−0.5943	0.2465	−2.4110
Trial	−0.1105	0.0448	−2.4700
1st syllable CVC : trisyllabic	−19.5494	8.7520	−2.2337
1st syllable CVCg : trisyllabic	−30.9267	10.7533	−2.8760
1st syllable CVV : trisyllabic	−24.8956	9.2714	−2.6852
1st syllable CVVC : trisyllabic	−12.5833	8.7740	−1.4342
1st syllable CVC : % non-modal vq.	−0.8640	0.2409	−3.5866
1st syllable CVCg : % non-modal vq.	−0.6014	0.2965	−2.0284
1st syllable CVV : % non-modal vq.	−0.7881	0.2494	−3.1604
1st syllable CVVC : % non-modal vq.	−1.0987	0.2395	−4.5881
Trisyllabic : % non-modal vq.	−0.8245	0.1409	−5.8535
Narrow focus : % non-modal vq.	−0.6027	0.1247	−4.8331
2nd syllable Q2 : % non-modal vq.	−0.7061	0.1502	−4.7024

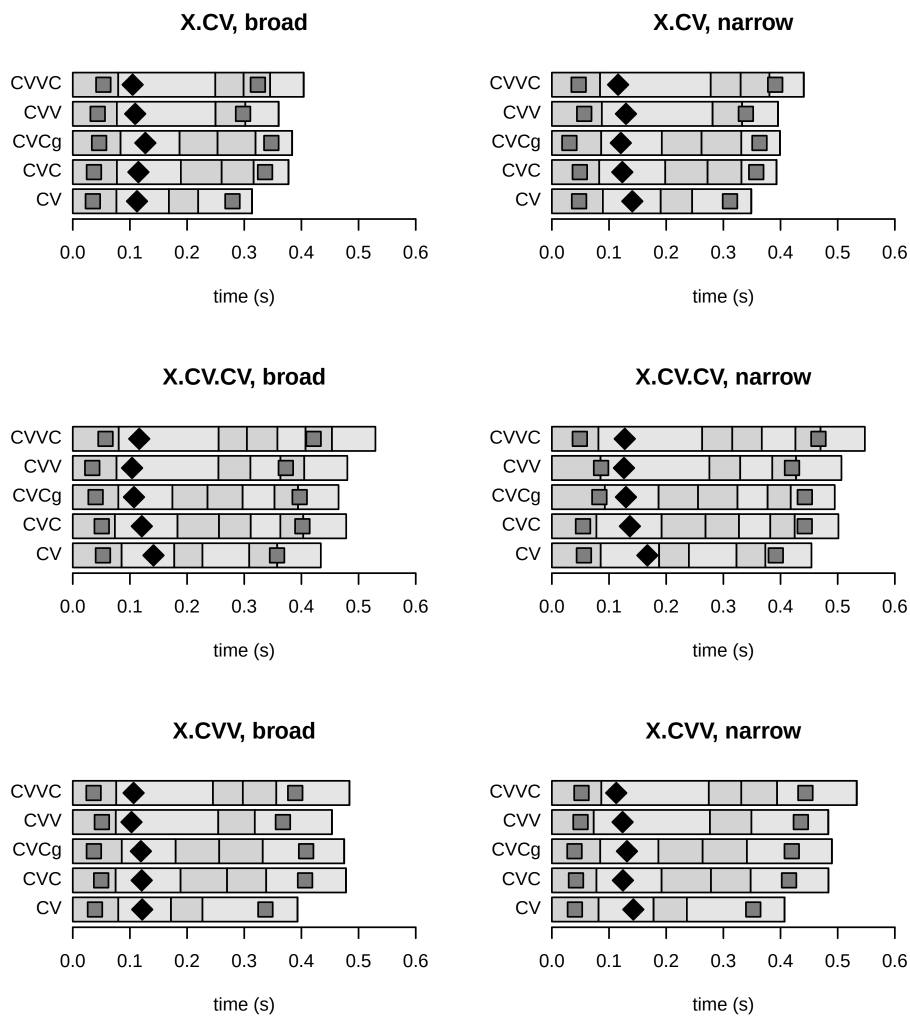

Figure 4.9: Average segment durations (in seconds) and timing of f_0 turning points for all experimental conditions in words without stretches of non-modal voice quality.

ment durations and timing for the three f_0 turning points, but it is based exclusively on the 1180 target words realised with modal voice throughout. Compared to Figure 4.4, mean times for H were very similar and also mean times for L1 did not differ systematically. By contrast, L2 alignment was measured consistently later in the absence of non-modal voice quality. This finding can also be expressed in terms of syllable alignment: In disyllabic words, L2 was measured on the second syllable for 90% of the completely modal words, but only for 69% of the data set on a whole. In trisyllabic words, 15% of L2 points appeared on the second syllable and 82% on the third syllable in the completely modal set, whereas 31% of L2 were on the second and 46% on the third syllable when partly non-modal target words were not eliminated. While above the beginning of the second syllable nucleus was determined as the best point of reference for L2 alignment in the data set on a whole, L2 was generally more closely aligned with the end of the second syllable nucleus in the completely modal words, with a median distance of -20 ms (vs. 57 ms from the beginning of the nucleus). However, note that even without the restrictions imposed by non-modal voice quality, mean L2 alignment as shown in Figure 4.9 did not gravitate around a certain fixed anchor point. The end of the second syllable nucleus constituted a better point of reference than other segmental landmarks, but mean L2 measurements in the longer essive and partitive words were at considerable distance from it. Figure 4.9 does not suggest the end of the second syllable target as an actual target of alignment. Instead, can most succinctly be described as follows: In the absence of non-modal voice quality, mean timing of L2 generally appeared during the word-final vowel or sometimes on the preceding consonant for short vowels. The median distance of L2 from the end of the words was -59 ms across all conditions.

4.4.5 Pitch scaling

The panels in Figure 4.10 show the effects of the experimental factors focus condition, first syllable structure and grammatical case (i.e. second syllable quantity and number of syllables) on pitch scaling. They depict mean values for low f_0 turning points before the beginning of target words (Lb), the words' first f_0 minima (L1), target words' maxima (H) and final minima (L2).

A clear difference between narrow and broad focus is visible and indeed the effect of focus condition was significant in all statistical models (see Table 4.16–4.19). Peaks and first low tones (with and without including Lb measurements) were higher in narrow focus. As both the figure and the model estimates demonstrate, the effect was largest for H tones and smallest for Lb tones. Substituting measured Lb for L1 values lowered the estimate for the effect of narrow focus, compare Tables 4.17 and 4.18. Also note that it caused values for the first minimum to be lower overall; the

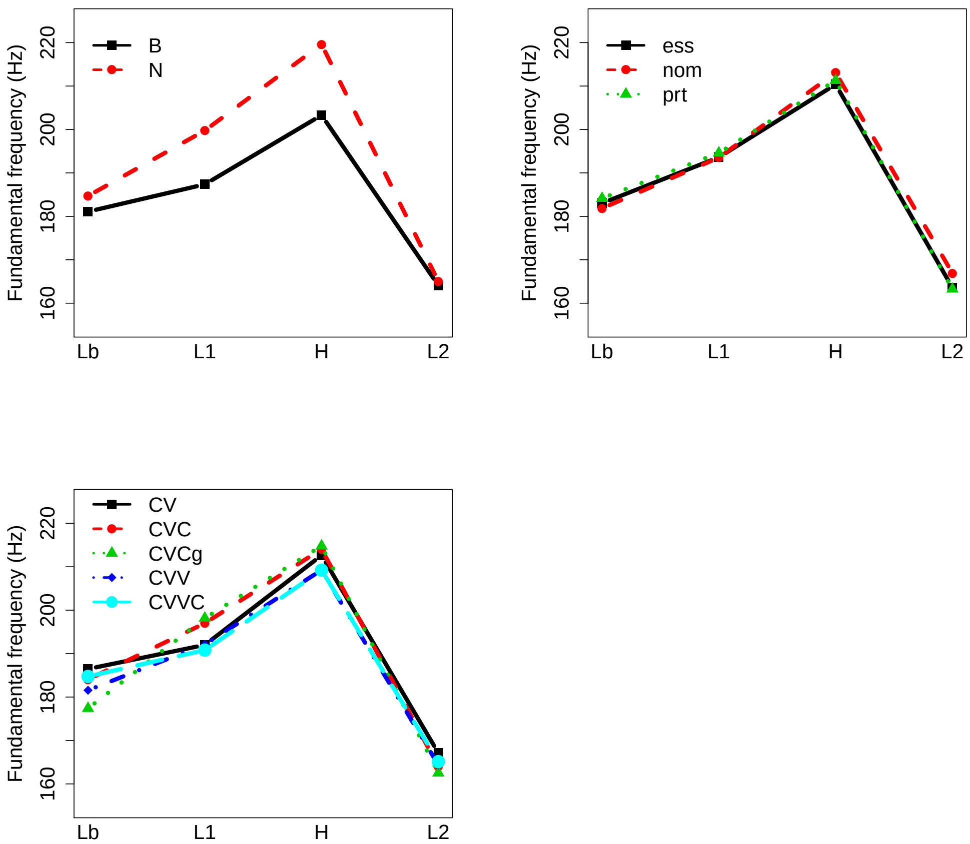

Figure 4.10: Mean values for pitch of Lb, L1, H and Lb in different experimental conditions (in Hz). Top left panel: All data. Top left panel: Focus condition; B = broad, N = narrow. Top right panel: Case; ess = essive, nom = nominative, prt = partitive. Bottom left panel: First syllable structure.

model in Table 4.18 indicates a lowered estimate for items where the editing process included checking whether the pitch rise began before the target word onset. For the scaling of L2, the 4.10 panel of Figure 4.10 show that the mean value was higher in narrow than in broad focus. Yet the model in Table 4.19 indicates that the main effect of narrow focus lead to lower pitch for L2. However, focus condition interacted with voice quality: Words with a larger proportion of non-modal voice realisation had a smaller effect of narrow focus, i.e. higher L2 tones in narrow focus. Additionally, there was a significant main effect of voice quality, with proportionally larger stretches of non-modal voice correlating with higher L2 pitch.

Table 4.16: Coefficients of the best linear mixed-effects model of the pitch peak (H, in st), with random by-item effects of number of syllables, focus condition and item list and with random by-subject effects of focus condition and trial. N = 2602.

	Estimate	Std. Error	t value
(Intercept)	24.6049	0.5124	48.0172
Trisyllabic	−0.1626	0.0749	−2.1703
Narrow focus	1.3253	0.1770	7.4894
List 2	1.0732	0.7249	1.4806
List 3	1.2731	0.7288	1.7469
Trial	−0.0041	0.0013	−3.2034
Male gender	−11.1200	0.9516	−11.6862

Table 4.17: Coefficients of the best linear mixed-effects model of f_0 minimum at the beginning of the target word (L1, in st), with random by-item effects of focus condition, trial and gender and with random by-subject effects of focus condition and trial. N = 1666.

	Estimate	Std. Error	t value
(Intercept)	24.3759	0.3448	70.7045
Narrow focus	1.1229	0.1525	7.3624
Trisyllabic	−0.1924	0.0652	−2.9517
Trial	−0.0052	0.0015	−3.6039
Male gender	−11.0422	0.8806	−12.5395

Effects of the other factors appear to be relatively small based on Figure 4.10, but they were nonetheless significant for some of the tones. In words with long second syllable vowel quantity, appearing in partitive, but not in nominative or essive words, pitch rises to the peak started higher and

Table 4.18: Coefficients of the best linear mixed-effects model of f_0 at the beginning of the rise to the peak (Lb or L1, in st), with random by-item effects of focus condition, item list and gender and with random by-subject effects of focus condition and trial. N = 1918.

	Estimate	Std. Error	t value
(Intercept)	23.0327	0.4865	47.3440
Narrow focus	0.9492	0.1277	7.4332
2nd sylllable Q2	0.2171	0.0741	2.9290
Checked for Lb	−0.5539	0.1514	−3.6577
List 2	1.0806	0.6664	1.6214
List 3	1.1624	0.6690	1.7376
Trial	−0.0045	0.0013	−3.6009
Male gender	−10.4905	0.8729	−12.0178

Table 4.19: Coefficients of the best linear mixed-effects model of f_0 minimum at the end of the target word (L2, in st), with random by-item effects of item list and with random by-subject effects of first syllable structure, second syllable quantity, focus condition and number of syllables. N = 2592.

	Estimate	Std. Error	t value
(Intercept)	21.2152	0.5623	37.7289
1st syllable CVC	−0.5243	0.1625	−3.2268
1st syllable CVC_g	−0.5147	0.1732	−2.9726
1st syllable CVV	−0.4657	0.1607	−2.8989
1st syllable CVVC	−0.4342	0.1633	−2.6594
Trisyllabic	−0.5093	0.1309	−3.8919
Narrow focus	−0.4224	0.0814	−5.1912
2nd syllable Q2	−0.5564	0.0811	−6.8582
% non-modal vq.	0.0210	0.0022	9.4348
List 2	1.0222	0.7441	1.3739
List3	1.8752	0.7249	2.5869
Male gender	−11.6245	0.9374	−12.4009
Narrow focus : % non-modal vq.	0.0190	0.0025	7.5493
Trisyllabic : % non-modal vq.	−0.0061	0.0027	−2.2591

falls ended at a lower level (see Tables 4.18 and 4.19). In trisyllabic words, i.e. those in essive case, peaks and word-internal minima were significantly lower than in disyllabic words, but no effect was found for first minima when taking realisations before the beginning of the word into account (see Table 4.18). For the scaling of L2, the number of syllables also interacted with the percentage of non-modal duration. In trisyllabic words the effect of proportionally large non-modal stretches was smaller than in disyllabic ones (see Table 4.19). The model of L2 scaling is also the only one exhibiting a significant effect of first syllable structure. Compared to CV.X words, pitch falls reached a lower level in words starting with all other initial syllable types.

Additionally, most models were improved by including the item list in the specification of main and / or random effects and, non-surprisingly, all measured pitch points were overall significantly lower for subjects who identified as men than for women.

4.5 Discussion

To a large extent, this chapter's findings confirmed those of the previous chapters and the hypotheses based on them. Additionally, the results added to and refined conclusions from Production Experiment I. Lastly, some findings regarding tonal analysis, focus marking and quantity or syllable structure effects differed from the outcome of the first experiment.

4.5.1 Tonal analysis and phrasing

The present chapter strengthened the analysis in terms of H_pL_p tones. Several findings confirmed the assumption that the target words' first minima measured as L1 did not correspond to phonological tonal targets. Often, L1 was not marked at all, because no seizable pitch rise preceded the peak within the target word. When present, L1 was invariably measured very close to the beginning of the word, without any influence of the experimental variables (see section 4.4.4.2). This is in line with an interpretation of L1 as part of an interpolation between L_p of the previous p-phrase and H, i.e. H_p of the p-phrase including the target word, as illustrated in Figure 4.11. When the editing process checked whether the pitch rise to the peak (H_p) already started before the target word onset, this was found to be the case for about 40% of the data even with the rather conservative method used (see section 4.4.3). Substituting these Lb measurements before the target word onset for the original the L1 measurements not only resulted in overall clearly earlier timing, but also in lower pitch, again affirming the interpolation analysis. Even with this substitution, no influence of the experimental variation was found on alignment. However, pitch of the first low tones was still affected by narrow focus on the target word (and its second syllable vowel quantity). It is not clear whether this was mostly a residue of remaining word-internal and thus interpolated L1 values or whether L_p of the preceding phrase was also affected.

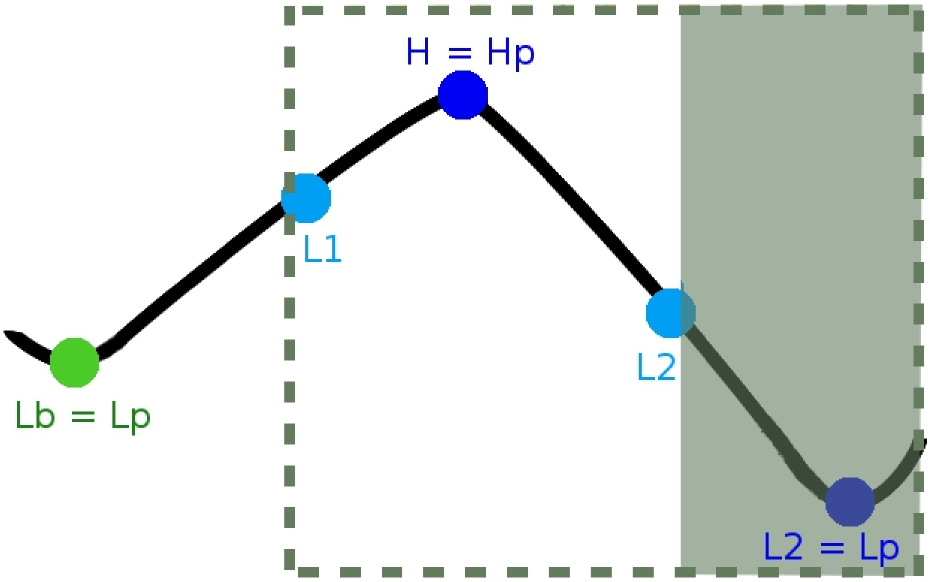

Figure 4.11: Stylised illustration of correspondence between phrase tones and f_0 turning points inside and outside of the target item (dotted box; grey area represents non-modal voice quality). Tones associated with the p-phrase containing the target word are rendered in blue, the L_p of the preceding p-phrase is given in green.

(26) Tonal targets associated with the p-phrase

$$\begin{array}{ccc} H_p & & L_p \\ | & & | \\ (& \ldots &)_p \end{array}$$

As stated in (26), this thesis assumes that the beginning of a p-phrase is associated with a high tone H_p and its end with a low tone L_p. In contrast to the finding of low turning points frequently preceding the target word, pitch peaks preceding the word onset or falls continuing into the following word were very rarely detected. Thus, the results confirmed the prediction in section 4.3.2 that pitch rises starting before the target words are more frequent than falls continuing after the target word and is in line with the postulation of p-phrases marked by H_pL_p tones. The observed distribution of low f_0 turning points before and after the target suggests that target words were usually phrased into p-phrases of their own or together with preceding words, yielding the tonal pattern schematically illustrated in Figure 4.11. This interpretation ties in with the absence of high turning points after the word and the rarity of high turning points before target onset likewise suggesting that target items constituted p-phrases of their own or were phrased together with preceding words, but attracted the H_p, as shown in (27a) and (27b), respectively.

(27) Assumed p-phrasing of target word (underlined)

a. Target word in own phrase:

$$\begin{array}{cc} H_p & L_p \\ | & | \\ \multicolumn{2}{c}{(\underline{\mathrm{PW}})_p} \end{array}$$

b. Target word phrased with preceding prosodic word(s):

$$\begin{array}{ccc} & H_p & L_p \\ & | & | \\ \multicolumn{3}{c}{(\mathrm{PW}\ \underline{\mathrm{PW}})_p} \end{array}$$

As illustrated in Figure 4.11, it is assumed that H_p tones corresponded to target word maxima H in the present data or, stated the other way around, that measured H tones were realisations of H_p tones. Maxima were usually localised early in the target word, as can also be deduced from the frequent absence of word-internally preceding minima L1.

A more minute investigation of peak timing revealed some variation, but no unified picture. The analyses showed two contradictory tendencies, which were not always significant. On the one hand, peaks were later in words starting with CV syllables than for words with longer and more complex first syllables. This was not a binary distinction between short and long first syllables, but a continuous trend towards later peaks in words with shorter first syllables and earlier peaks in those with longer first syllables. For the subset of data consisting of direct minimal pairs with the shortest and longest initial syllables, CV and CVVC, the difference was significant. On the other hand, the effects of the factors number of syllables, second syllable quantity and focus condition showed a unified tendency towards later peaks in longer words. The most stable of these effects, appearing as significant in most subset models, was that of later peaks in narrow focus, which at the same time showed a slightly narrower distribution. In sum, the evidence for the significance of these two contradictory tendencies was mixed and the differences between the conditions were quite small compared to the variation inside the individual conditions. It does thus not provide grounds for rejecting the uniformity hypothesis nor for assuming phonological relevance of peak timing differences. Regardless of minor variations, alignment of pitch maxima was overwhelmingly detected early in the target words. Altogether 96% of H tones occurred on the first syllable vowel or the preceding onset consonant, in consonance with the interpretation of maxima as realisations of H_p tones associated with the beginning of p-phrases.

The segmental alignment of L2 varied much more. The findings of section 4.4.4.3 suggest that its timing was mainly determined by two factors: voice quality and word duration governed by focus condition and syllable structure (i.e. number of syllables, first syllable type and second syllable quantity). As illustrated in Figure 4.11, pitch falls from the target word peaks were frequently cut short by non-modal voice quality: Timing was significantly earlier and pitch measured for L2 was significantly higher in words with a larger percentage of non-modal duration (see also section 4.4.5). In line with the hypothesis of p-phrase final L_p, L2 was measured later and with lower pitch in conditions causing longer word duration, i.e. in narrow focus vs. broad focus condition, in trisyllabic vs. disyllabic words and in words with longer or more complex first or second syllables vs. those with first or second CV syllables. When pitch falls were not restricted by non-modal voice, they generally ended close to the target word offsets. For example in trisyllabic words with modal voice throughout, over 80% of L2 were measured on the final syllable, further supporting the association of L_p with the end of p-phrases.

Apparent effects of gender on alignment were most probably due to the fact that only three participants identified as male, two of whom (s12 and s21) were assigned to item list 3 (see Appendix 4.2.3). Since this list was also edited last, all items on this list were checked for the presence of f_0 turning points outside the target words. This affected the measured timing of tones, but did not necessarily reflect any real differences.

4.5.2 Focus marking

Effects of narrow focus were largely the same as summarised for Production Experiment I in Figure 2.23: Compared to broad focus, narrow focus words had a larger pitch range with higher H_p peaks followed by falls to lower L_p tones (see section 4.4.5 above). They also showed longer durations (see section 4.4.1) and more frequent non-modal voice quality on the second (and, when present, third) syllable (section 4.4.2). Pertaining to voice quality, the current data clarified the conclusions drawn from Production Experiment I. It revealed that also in trisyllabic words, the narrow focus words had an increased frequency of non-modal realisations already for the second syllables. The narrow focus effect can thus be localised relative to the beginning of the respective word (or phrase) and not primarily relative its end, although section 4.4.2 found that the distance from the end of the word had an additional influence.

One of the prosodic focus marking methods discovered in Chapter 2 and summarised in Figure 2.23 was not reproduced in the current chapter's materials. Section 4.4.2 found no difference between focus conditions in the frequency of pauses before or after the target word. However, this particular strategy of focus marking was only investigated for the non-final constituents subject and verb in Chapter 2, not for the objects, which were more directly comparable with the target words in the current materials.

4.5.3 Quantity and syllable structure

Effects of quantity and more generally syllable structure were also largely as predicted in section 4.3.2. Namely, they can be captured by stating that target words with more and / or phonologically long segments showed longer duration. Word duration was shorter for words beginning with CV syllables than for those beginning with any other syllable types and the mean durations showed the predicted hierarchy CV < CVC, CVC_g, CVV < CVVC. Likewise, words with second syllable Q2 vowels hand longer duration than those with Q1 vowels and trisyllabic words were longer than disyllabic ones. These data can also be taken as support for viewing the mora as a timing unit, since all bimoraic syllable types showed very similar durations. Also, monomoraic syllables were significantly shorter than bimoraic syllables, which were in turn clearly shorter than trimoraic ones. This ranking thus supports a tertiary distinction between light, heavy and superheavy or short, long and overlong syllables. Although further research should be conducted to verify the three-way distinction for non-initial syllables, note that evidence for the mora as a timing unit in Finnish is also presented by O'Dell et al. (2007). In their material from spontaneous conversational speech, mora count had a significant effect on pause group duration (the duration of a chunk of speech delimited by preceding and following pauses).

The present chapter also found effects of the factor quantity on other phonetic variables. These are explained as by-products of duration effects here. This was already illustrated with regard to alignment in section 4.5.1 above. Likewise, the findings regarding syllable structure and voice quality can be modelled by assuming a word-initial stretch of modal duration with an absolutely fixed minimal duration, as illustrated in Figure 4.12. This initial modal stretch provided space for realising H_p and a sufficient pitch fall to indicate the realisation of L_p. After this initial modal interval, speakers frequently switched to non-modal voice which is interpreted as a phrase-finality marker here. The assumption of a minimal modal duration fixed in absolute terms would need to be verified directly. It does however seem functionally plausible and accounts well for the present chapter's findings: Like in the data from Production Experiment I, the frequency of non-modal realisations in first syllables was lower when these first syllables were CV than for the other syllable types, confirming the hypothesis stated in section 4.3.2. This makes sense when assuming that the fixed modal duration was more likely to completely extend over the shorter CV syllables than over the longer syllable types (see the difference between the top two bars in Figure 4.12). Also, non-modal realisations of second syllables were more frequent in partitive case, where second syllable vowels were long (Q2) instead of short (Q1). Again, an explanation in terms of an absolute minimal modal duration is plausible, in that this stretch was less likely to cover the second in addition to the first syllable when the second syllable was longer (see the contrast between the bottom two bars in Figure 4.12).

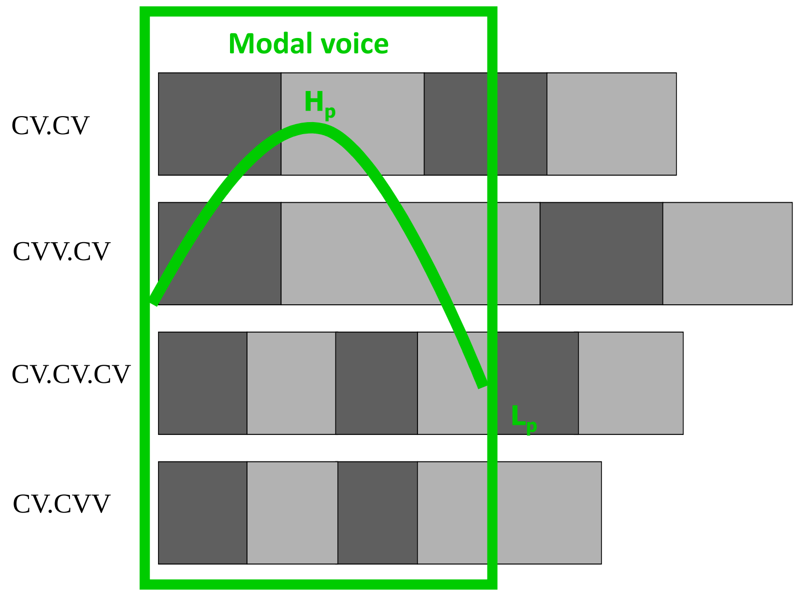

Figure 4.12: Stylised illustration of the effects of a modal voice quality stretch with fixed duration for words with different syllable structures. Dark grey bars symbolise consonant durations, light grey bar vowel durations.

One can also ask whether narrow focus effects for example on pitch scaling can be explained primarily in terms of longer word durations. However, such a view does not seem appropriate, since word duration showed significant effects of syllable structure as well as focus condition, but while pitch scaling was consistently influenced by focus condition, effects of syllable structure were much less consistent (see section 4.4.5).

4.6 Summary

This chapter presented a production study investigating f_0, duration, occurrence of pauses and voice quality for narrow and broad focus words with varying syllable structures in pre-final position. While effects of focus condition were found for f_0, duration and voice quality, findings regarding syllable structure—i.e. differences in first and second syllable quantity and complexity and in number of syllables—were explained largely in terms of effects on duration.

Importantly, the current chapter confirmed the analysis of pitch contours as largely shaped by phrase tones associated p-phrases. It investigated the alignment of these tones, $\mathrm{H_p}$ and $\mathrm{L_p}$, more precisely and concluded that they are associated with the beginning and end of the p-phrase, respectively. While minima generally appeared close to the target word ends—unless non-modal voice quality curtailed the (measurable) pitch falls—pitch maxima were overwhelmingly observed on the target words' first syllable nuclei. The chapter thus re-iterated the finding that there is no conclusive evidence for phonologically relevant differences in peak timing.

Chapter 5

Perception Experiment: The role of the pitch contour in the perception of compounds and noun phrases

5.1 Introduction

This chapter concentrates on the second research question of this thesis, namely the question of the shape and appropriate tonal specification of intonation contours (see (10b) in section 1.7). The thesis argues that Finnish sentence intonation is best accounted for in terms of phrase tones (see Féry, 2010, on phrase languages). Chapters 2 and 3 suggested to model the pitch range effects of focus as an adjustment of two p-phrase tones, H_pL_p, and Chapter 4 further investigated their alignment. The current chapter examines the postulation of these two tones with evidence from language processing. Specifically, it reports an experiment that investigated which pitch movements need to be realised on a word so that a listener perceives it as having a contour of its own. That is, according to the current analysis, it tested which movements induced the perception of the respective word as a p-phrase of its own. However, this question is relevant not only within the current phrasal analysis, but also with respect to previous accounts in terms of accents. Phonological accounts of Finnish intonation agree in describing accents that are of the same type or shape, with the only exception of a few marginal cases (see Välimaa-Blum, 1993; Iivonen, 1998; Suomi et al., 2008), but different specifications of this uniform accent have been suggested. Whereas Välimaa-Blum (1993) finds L+H* and rarely L*+H accents in her material, Suomi et al. (2008) argue that Finnish accents need to be specified as LHL, explicitly ruling out a description with only two tonal targets per accent, e.g. L+H*, H*+L etc.

The experiment discussed in this chapter addressed the question of tonal specification from the point of view of perception. It employed segmentally ambiguous items like *märkäpuku* ‘wetsuit’ that are only identified by prosody as a two-word compound or noun phrase (i.e. *märkäpuku* ‘wetsuit’ vs. *märkä puku* ‘wet suit’). As shown by Niemi (1984), such minimal pairs are distinguished by duration, f_0 and intensity in Finnish, with intensity constituting a less important cue than duration and f_0 (see Niemi, 1984, pp. 92–93). This chapter will concentrate on pitch cues to the distinction. Figure 5.1 gives a schematic illustration of the different f_0 contours as observed by Niemi (1984), who compared pronunciations by native Finnish and American English speakers. Notice that while the English speakers produced realisations with only one accent, differentiating the two structures through accent location (see e.g. the English compound stress rule in Chomsky and Halle, 1968), Finnish speakers realised the compounds with one shared tonal contour and the noun phrase versions with two rise-falls. The same difference for contours of Finnish noun phrases and compounds appeared in the present materials, i.e. compounds displayed one shared contour, while noun phrases showed two separate rise-falls. However, in contrast to Niemi’s description, compounds also started with a clear pitch rise followed by a fall, see Figures 5.2 and 5.3.

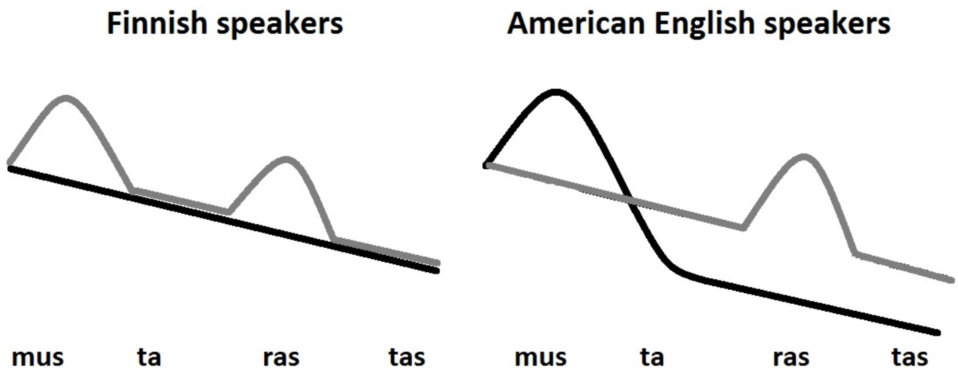

Figure 5.1: Schematic illustration of f_0 contours for realisations of the item *mustarastas* as a compound meaning ‘blackbird’ (black lines) and as a noun phrase meaning ‘black thrush’ (grey lines), as spoken by Finnish and American English speakers, according to Niemi (1984, p. 76).

This chapter analyses the distinction as two-word compounds being two prosodic words united in one p-phrase and two-word noun phrases as being two prosodic words in separate p-phrases (an accent-based account could state that noun phrases carry two accents, while compounds are realised with only one). Building on this assumption, a perception experiment used items with manipulated f_0 contours to test which f_0 movements induced the perception of a separate phrase (or accent) on the second part and thus of a noun phrase. Specifically, the study investigated whether three tonal

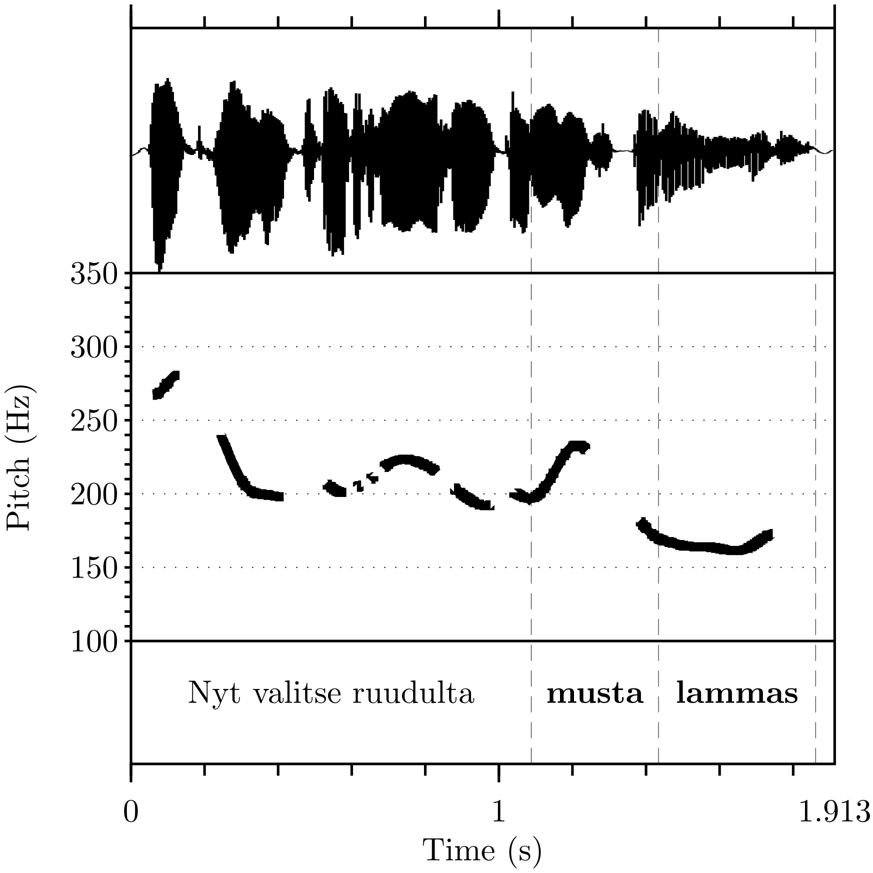

Figure 5.2: Pitch contour of the sentence *Nyt valitse ruudulta mustalammas* 'Now select the black sheep from the screen'. The compound *mustalammas* 'black sheep' refers to a person who ignores traditions and sticks out from a group.

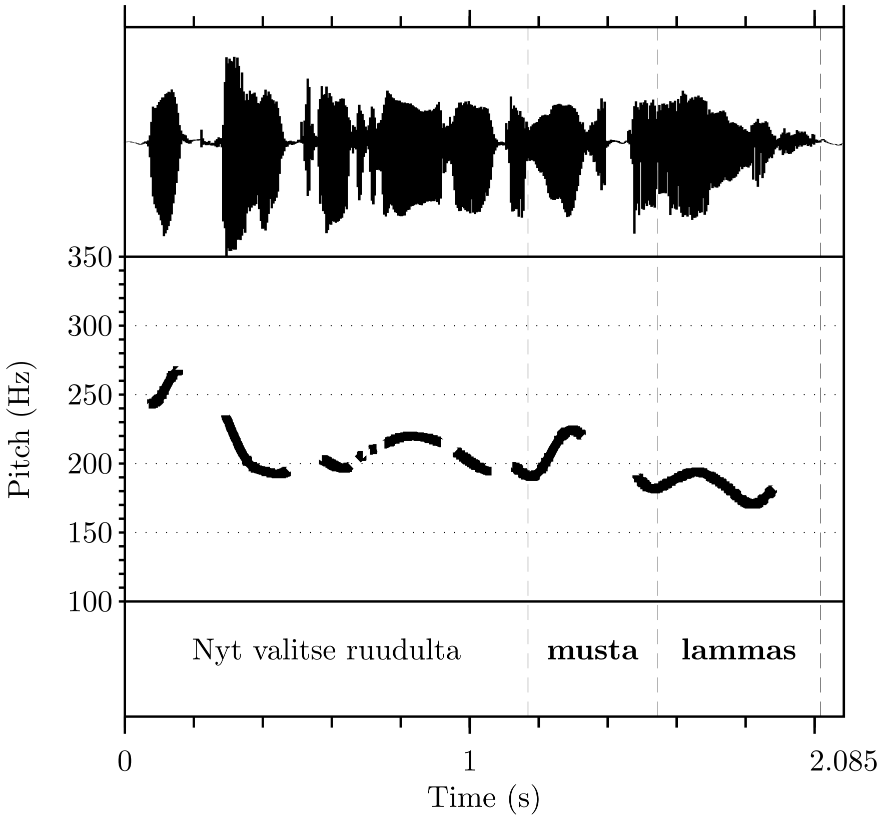

Figure 5.3: Pitch contour of the sentence *Nyt valitse ruudulta musta lammas* 'Now select the black sheep from the screen'. The noun phrase *musta lammas* 'black sheep' refers to an animal.

targets LHL, i.e. a complete rise-fall, need to be realised for native speakers to perceive a separate p-phrase or whether 'incomplete' contours corresponding to bitonal specifications (rises or falls) have the same effect.

5.2 Methods

The current experiment investigated the effects of pitch manipulations on speech processing with a two-alternative forced choice task. Participants were not asked which word they heard, but had to choose a picture according to an auditory instruction which included a manipulated target item. They were thus presented with a visual and an auditory stimulus in parallel, as described in more detail below.

5.2.1 Target and filler items

Items representing compound / noun phrase minimal pairs of the type 'wet-suit' vs. 'wet suit' were used. For the picture selection task to capture the structural difference, referents of both meanings needed to be depictable and sufficiently different from each other. This excluded pairs like *talonmies* 'janitor' vs. *talon mies* 'the man of the house', as well as a translation of the English example 'blackbird', since the meaning of the noun phrase equivalent *musta rastas* 'black thrush' is much more specific (there is no other black thrush apart from the blackbird).

Thirty-four suitable minimal pairs were found, for which a 29-year old university student from Southern Finland confirmed that she was familiar with the compound meanings. In addition to these target items, which were segmentally ambiguous between noun phrases and compounds, 34 filler items pairs were also selected. These consisted of an unambiguous noun phrase and an unambiguous compound. To make the picture selection non-trivial also for the filler trials, the members of the filler item pairs were phonetically similar, as in *mustekala* 'squid' vs. *musta kala* 'black fish', or visually much alike, as for the pair *raitiovaunu* 'tram car' vs. *vihreä juna* 'green train'. Furthermore, to make item and filler sets more similar, each filler item pair was matched to a target item pair by compound lemma frequency per million, calculated on the basis of the Karjalainen lexical database.[1] Target and filler compound frequency were not always exactly the same, but very similar, as can be seen in Appendix L listing all target and filler items. Filler compounds were also overall roughly matched to the targets in terms of word length, so

[1] The Karjalainen database is based on the annual volumes of the Finnish newspaper *Karjalainen* for the years 1991 to 1997. It contains 34.5 million word tokens and was compiled by Jussi Niemi and his colleagues at the University of Eastern Finland (formerly University of Joensuu), SGML form created at the Department of General Linguistics of the University of Helsinki. The corpus is available online through CSC, (`http://www.csc.fi/`).

that all target and filler compounds were between 7 and 16 letters long (in fact, only one of the targets, *harakanvarvas*, was longer than 13 letters).[2]

Pictures of the filler items were shown six times per session, so that participants would encounter them exactly as often as the pictures of the target item referents. In contrast to the target items, they were accompanied by an auditory stimulus that unambiguously identified either the compound or the noun phrase referent as the correct choice, inducing both choices equally often. The auditory stimuli were thus prepared slightly differently for target and filler items.

5.2.2 Auditory stimuli

All items, targets as well as fillers, were recorded in the carrier sentence *Nyt valitse ruudulta* ____ 'Now select ____ from the screen'. They were spoken by a 26-year old native speaker of Finnish from Helsinki. The speaker, an advanced student of Phonetics, was told that the utterances would be used as prompt sentences in an experiment, but was not informed of the objective of the study. In the first recording session, she read 108 sentences from a list containing compound target items, filler compounds, filler noun phrases and noun phrase versions of the target items, repeating each sentence at least three times. Figures 5.2 and 5.3 above show examples of these original recordings. In a second recording session, 14 of the filler items and the practice items were recorded separately.

In a next step, I chose one realisation for each of the items, confirming its naturalness and suitability as a neutral realisation of the desired word or phrase with a native speaker. The chosen filler utterances were resynthesised with PSOLA implemented in Praat (Boersma and Weenink, 2010) as close approximations of the original pitch contours. From the chosen realisations of the target utterances, stimuli were created in the Praat programme with the following procedure:

First, a Praat manipulation object was created from the compound version of each item (e.g. Figure 5.2). The original f_0 contour was modelled with maximally five pitch points per word. Three f_0 turning points—L1, H and L2 as measured for the production data in Chapters 2 and 4—were marked for measurement in the first part of the target compound, e.g. for *märkä* in *märkäpuku* 'wetsuit'. These measurement points were always included in the pitch contour of the manipulation object. Additionally, the beginning and end of the compound, as well as the boundary between its two parts were marked. The segmentation criteria were the same as those used for Production Experiment II, with the exception that the end of the last word was not marked on the basis of the first formant, but based on the oscillogram to be sure not to cut any part of the word off.

[2] This definition of word length was of course due to using a corpus of written language for finding filler compounds matching the targets items in frequency.

Second, the corresponding word boundaries and the same three f_0 turning points were also marked in the recordings of the noun phrase versions, for both words of the target noun phrase.

Third, the manipulated sound files, which were used as prompts in the perception experiment, were created. The compound versions were chosen as the source utterances for all manipulations so that a possible bias of temporal cues would be conservative. In other words, since the question of the experiment was whether certain pitch movements can induce the perception of a separate tonal contour on the second part of a possible compound and thus of the target item as a noun phrase, it was necessary to ensure that a noun phrase classification of the manipulated stimulus was truly induced by the f_0 contour it carried and not by temporal or other residue in the utterance.

Six different manipulations were created for each item, as schematically illustrated in Figure 5.4. These contours are in the following referred to as 'flat', 'rise-fall', 'rise', 'fall', 'high fall' and 'high flat' according to the pitch contour present on the target word, i.e. the second part of the ambiguous compound / noun phrase stretch (e.g. *puku* 'suit' in *märkäpuku*). The pitch of the first part (e.g. *märkä* 'wet' in *märkäpuku*) was not manipulated for the patterns 'flat', 'rise-fall', 'rise', 'fall', but the patterns 'high fall' and 'high flat' lacked the pitch fall at the end of the first part. Figure 5.4 shows both parts of the compound / noun phrase for clarity, because the first part was relevant in determining the pitch scaling and timing of the second part for all manipulated contours.

In most of the manipulations, only the sentence-final second prosodic word of the compound / noun phrase was thus manipulated (e.g. *puku* in *märkäpuku*), while the rest of the utterance's f_0 contour as created in the first step remained intact. From the final word, all f_0 points were removed and new tonal contour was imposed on it. The six manipulations were created as follows:

1. **'Flat' pattern** The last pitch point at the end of the first part of the target compound was repeated at the end of the second part (1 ms before the word boundary). The pitch on the second word thus remained completely flat at a low level.

2. **'Rise-fall' pattern** Like for the 'flat' pattern, the last pitch point of the first compound part was copied to the end of the second part. Additionally, a peak was inserted on the manipulated word so that it carried a rise-fall pattern.

   The timing and f_0 of this peak were determined based on the noun phrase recording of the same item to achieve a pitch movement as similar to naturally occurring shapes as possible. The temporal distance of the added peak from the beginning of the manipulated word was

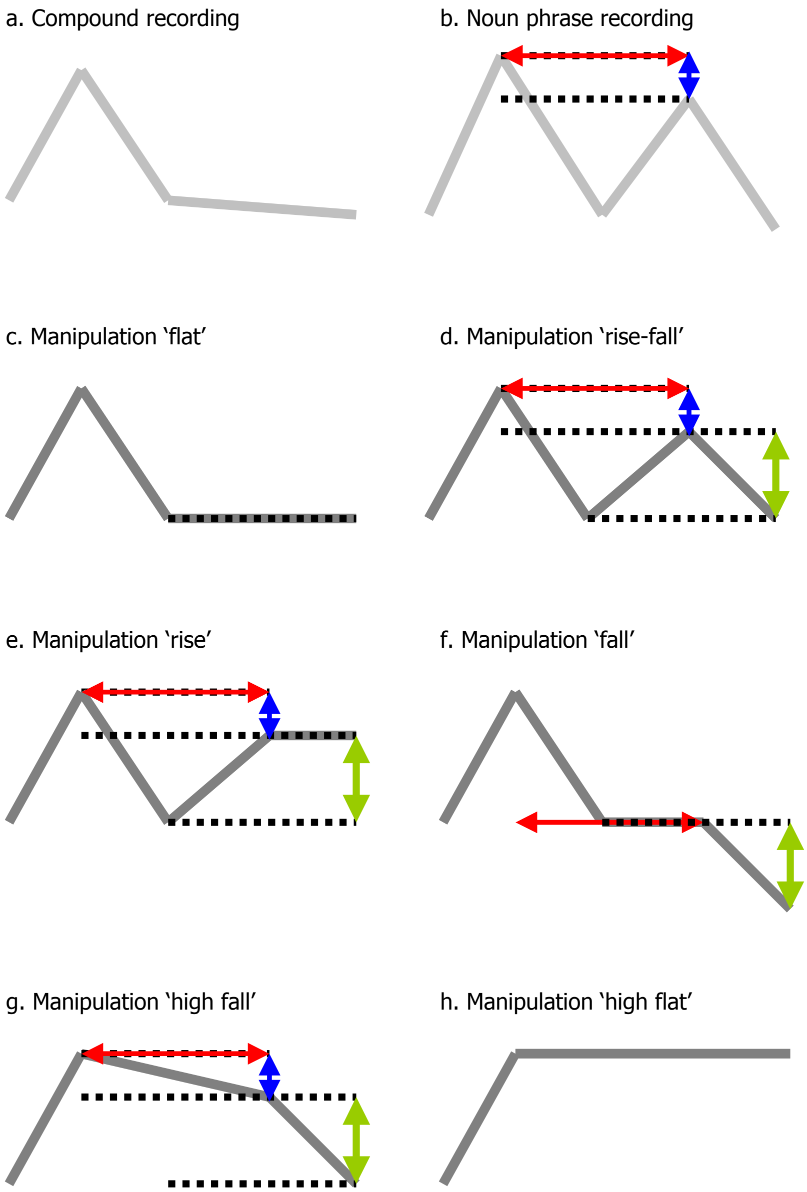

Figure 5.4: Schematic illustration of pitch patterns on the target compound / noun phrase.

the same as measured for the peak realised on the second word of the noun phrase in the corresponding recording of the same lexical item. In terms of f_0, the distance to the preceding peak was also the same as in the noun phrase version of the same sentence. That is, the difference between the two peaks of the target item in the noun phrase recording (in Hz) was subtracted from the peak value of the first compound part to calculate the f_0 of the peak added on the second part of the compound.

For three items—*harakanvarvas, juoponnappi* and *verenpisara*—the peak resulting from this procedure was quite low. Therefore, the peak was set to exactly the same value as that on the second word of the noun phrase in the corresponding recording of these items.

3. **'Rise' pattern** For this manipulation, a high pitch point was added to the second part of the compound with the same timing and f_0 as the peak in the 'rise-fall' pattern. No low pitch point was added at the end of the word, however, so that the pitch on the word first rose to this point and then remained flat on a high level.

4. **'Fall' pattern** This pattern only included a fall on the second part of the compound, but no rise. The beginning of the fall was timed in the same way as the peak for the 'rise-fall' and the end of the rise in the 'rise' pattern. The f_0 at the beginning of the fall was the same as the last f_0 value in the first part of the compound (the same value repeated at the end for the 'rise-fall' and the 'rise' pattern). The size of the pitch fall was the same as in the 'rise-fall' manipulation of the same item.

5. **'High fall' pattern** A second manipulation with a falling contour was created. This pattern included a fall that was completely identical to the one occurring in the 'rise-fall' pattern. To achieve this without a rise up to a peak, all pitch points between the start of the fall and the peak of the first part of the target compound were deleted, producing a high, more or less tilted plateau. In contrast to the first four manipulations, the contour of the first part of the compound was thus not identical to the source utterance.

6. **'High flat' pattern** The materials included another manipulation in which the pitch contour of the first compound part was changed for comparability with the 'high fall' pattern. Here, all pitch points after the peak of the first compound part were deleted, forming a completely flat high plateau over the rest of the target until the end of the utterance.

Pitch was interpolated quadratically to give smooth contours before resynthesising with the PSOLA method for both target and filler stimulus sounds.

5.2.3 Procedure

Data collection took place in the sound-proof recording booth of the Faculty of Behavioural Sciences at the University of Helsinki in November 2011. Stimulus presentation and data collection was done using E-prime 1.1 and an E-Prime Serial Response Box, collecting response time and choice. Participants were told that two pictures would appear on the screen and their task would be to select the picture on the left or on the right by pressing the corresponding button according to instructions from headphones. They were asked to respond as fast, but accurately as possible.

Each experimental session started with six practice trials that did not occur again later in the session. The same frame sentence and the same method of sound resynthesis was used as for the fillers. The target items were three compounds (*maalivahti* 'goalkeeper', *ydinvoimala* 'nuclear power plant', *homekoira* 'mould detection dog') and three noun phrases (*äidin saappaat* 'mother's boots', *kitaran kaula* 'guitar neck', *lehmän sorkka* 'a cow's hoof'). Pictures of these were contrasted with semantically and / or optically similar pictures (indicating a forward player, a factory, mould cheese, a pair of sandals, a guitar body and a horse's hoof).

The experimental session proper consisted of two parts: Participants were first presented with all experimental items in the pitch conditions 'flat', 'rise-fall', 'rise' and 'fall', interspersed with two repetitions of the filler items in the noun phrase version and two repetitions of the fillers indicating the compound as the correct choice (altogether 272 trials). Second, after a break, they encountered all experimental items in the pitch conditions 'high flat' and 'high fall' together with another two repetitions of the fillers, once with the noun phrase and once with the compound as the correct choice (136 trials). The intention in splitting the materials up in this way was to separate the possibly confusing patterns with manipulations affecting the first part of the compound/ noun phrase from the rest of the materials.

Four different trial lists were used. The order of the practice items was randomized for each list. Within the two parts of the experiment proper, the order of the items was pseudo-randomized to ensure that the subject would not encounter the same lexical item twice in a row. Appendix K records the list used for each of the participants together with other background information.

The chain of events within a trial was the same for the two parts of the experimental session and for the practice trials. First, a fixation cross was displayed in the middle of the screen for 1000 ms. Second, the stimulus slide appeared, showing the two pictures representing the noun phrase and the compound referent of the item pair surrounded by a white frame. All pictures were landscape format photographs of the same size. For some pairs, an arrow was used in both pictures for clarification. To avoid priming, the location of the pictures was balanced so that the photograph showing the compound referent was always displayed on the left for half the items and on the right for the other half (in target, filler and practice trials alike).

The auditory stimulus was presented at the same time over Sennheiser HD 515 headphones. The onset of the soundfile was time-aligned with the appearance of the visual stimuli. Thus, the pictures were visible to the participant between 1338 ms and 1609 ms before the onset of the target word, depending on the articulation speed of the individual frame sentence. Response was enabled from the appearance of the slide. The pictures remained visible also after the end of the sound file, staying on the screen until the subject responded, but maximally for 4000 ms altogether. If a participant failed to respond within that time, the programme automatically proceeded to the next trial.

5.2.4 Participants

The study was carried out with 28 participants, but I chose a wrong experiment file to run for the 20th subject, so that the session could not be completed and one of the four trial orders was thus executed only six times.

The remaining 27 participants (20 female) were between 18 and 47 years old (mean age 25.11), students and had all attended primary school in the Helsinki metropolitan area. Two of them (subject 01 and 10) had already participated in Production Experiment I. Subject 03 identified as bilingual with Finnish clearly being her stronger language, her other language being French. Another participant, subject 17, reported worse hearing on his left ear, but no effect on his overall hearing ability. Since the prompts were mono sounds and input to both ears was identical, his performance should not have been influenced. None of the other participants reported any hearing loss. Further information on the subjects' background can be found in Appendix K. A few subjects remarked after the session that they did not know some of the lexical items. These cases are also noted in Appendix K.

All participants were reimbursed for their time with a shopping voucher.

5.2.5 Data editing and analysis

E-prime logged reaction times from the onset of the simultaneous auditory and visual stimulus presentation. To arrive at the reaction time relative to the onset of the target word (the second part of the compound / noun phrase), the distance between the target word onset and the beginning of the stimulus sound was subtracted from the reaction time registered by E-prime. All values were measured and calculated with an accuracy of one millisecond (ms).

Altogether 5412 responses and reaction times were collected from the target trials with the experimental items. The missing 96 data points, 2% of the 5508 trials run (34 lexical items $\times$ 6 pitch contour conditions $\times$ 27 participants) are due to participants failing to respond to the stimuli before the presentation automatically proceeded to the next trial.

However, the obtained responses included a few—eight—cases where participants had pressed the button before even hearing the target word, as the left panel of Figure 5.5 shows. Since these and other extremely early responses can be assumed to be erroneous, they should of course not be considered in the analysis. Therefore, responses earlier than 200 ms after the target onset were removed. This concerned 49 data points (1% of the collected data), while a cut-off at 100 ms would have lead to discarding 25 responses (less than half a percent). As a result, 5363 observations were retained for analysis.

In contrast, late responses were not treated as outliers, since no principled distinction between genuine and spurious long reaction times could be made (see also Ratcliff, 1993, who shows that long reaction time outliers are generally more difficult to discover than short ones). In the edited data set, all but 131 responses (2%) only occurred after the offset of the target word; however, the experimental procedure automatically proceeding to the next trial introduced an upper boundary.

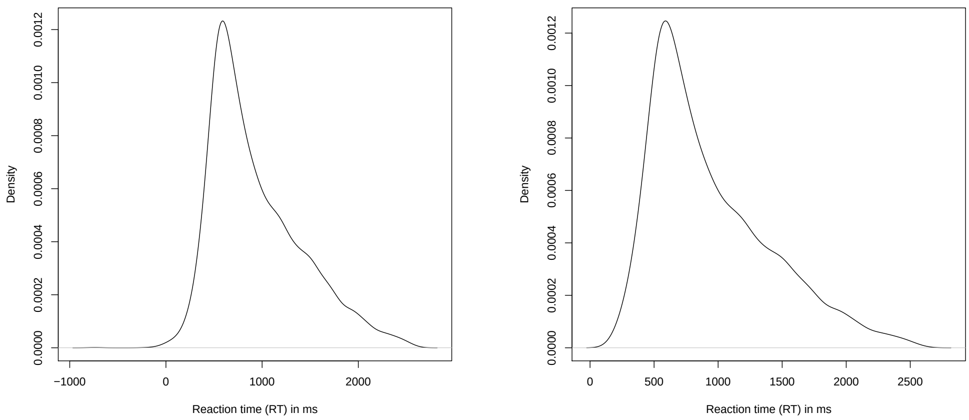

Figure 5.5: Density distribution of reaction times for the unedited data set (left) and with reaction times below 200ms removed (right).

For statistical analyses, reaction times were log-transformed. This resulted in a close-to normal distribution, as can be seen in Figure 5.6. All statistical analyses were done by fitting linear mixed-effects models as implemented in R to the data (Baayen, 2008; R Development Core Team, 2010). For the choice between noun phrase and compound interpretation, these models were binomial and the function automatically calculated p-values based on the z-score. This was not the case for the models of reaction time. Also, for models with complex random factors like the ones used here, the calculation of p-values through Markov chain Monte Carlo simulation (Baayen, 2008; R Development Core Team, 2010) is not implemented. Since the data set is relatively large, significances were therefore estimated on the basis of t-values. Effects were assumed to be significant when t > |2|.

When it was necessary to raise the number of iterations to make a model converge, the maximum number of iterations was set to 1000. In determining the best-fitting models for each measure and (sub)set of data with an ANOVA comparison, the model with the larger log likelihood was chosen in case of significant differences, even if it also had larger values for the Bayesian information criterion (BIC) or Akaike information criterion (AIC). This means that the chosen models might constitute an overfit, possibly making effects seem weaker than they are. This decision was made to lower the risk of type I errors.

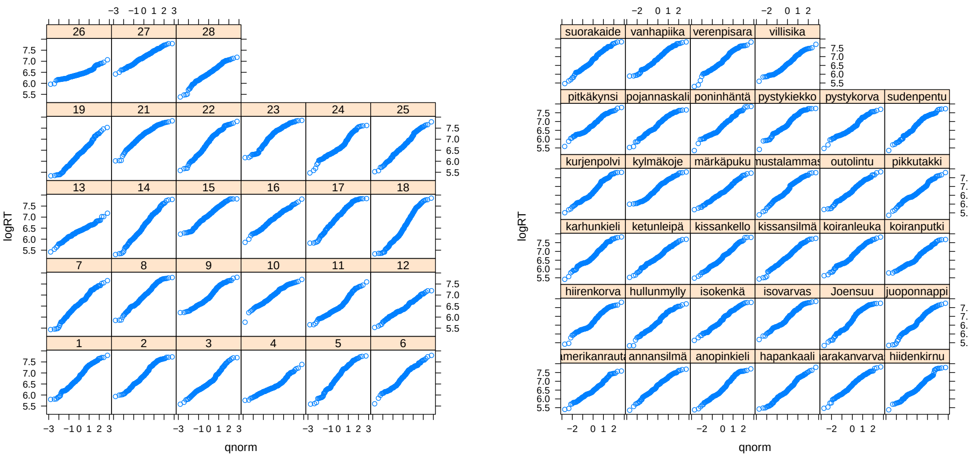

Figure 5.6: Quantile-quantile plots for log-transformed reaction times by subject (left) and by item (right).

5.3 Hypotheses

The study is based on the pitch differences between compounds and noun phrases, namely that—as described in the introduction—two-word compounds are realised as one p-phrase spanned by the realisation of H_pL_p tones, while two-word noun phrases are regularly realised as two p-phrases with separate contours on both (or, in accent-based approaches, compounds bear one initial accent, noun phrases an accent on each of their two parts).

Before introducing the hypotheses regarding the results of the experiment, a general hypothesis bearing on experimental design should be made explicit: This study assumes that it is indeed possible to create a stimulus that will be perceived as a noun phrase from a compound recording purely by manipulating the pitch contour. Preliminary tests with native speakers successfully elicited noun phrase perceptions in this way, whereas it seemed impossible to turn a noun phrase recording into something that speakers perceive as a compound by manipulating only the f_0 contour.

Regarding the participants' predicted reactions to the experimental materials, a general hypothesis is that there should be differences between the six manipulation patterns used for each item. The 'flat' contour is clearly expected to lead to the highest percentage of compound responses, since it is most similar to the pitch contour that naturally occurring phrase-final compounds carry. Conversely, the 'rise-fall' contour corresponding to natural realisations of noun phrases should lead to a significantly higher percentage of noun phrase responses.

Table 5.1: Summary of predictions based on three different accounts of Finnish intonation.

	Predictions		
	$\mathbf{H_pL_p}$	**L+H* / L*+H**	**LHL**
More NP responses ↑	Rise-fall	Rise-fall	Rise-falls
	High fall	Rise	
			Rise
	Rise	Fall	Fall
	Fall	High fall	High fall
	High flat	High flat	High flat
Less NP responses ↓	Flat	Flat	Flat

The prognosis for the other contours depends on which account of Finnish intonation one follows, cf. the overview in Table 5.1. A schematic illustration of the predictions appears in Figure 5.7. It shows annotations of the six contours in terms of H_pL_p tones as suggested in this thesis, in terms of Välimaa-Blum's (1993) L+H* accents and in terms of LHL accents following Suomi et al. (2008). These annotations mark tones corresponding to minima or maxima in the contours in black, while missing or unrealised tones are marked in red. Light blue boxes highlight annotations of complete tonal contours on the second prosodic word of the target item according to the respective analysis. These patterns are expected to induce significantly more noun phrase responses. Thus, all three annotations agree in marking the second prosodic word as having a complete tonal contour of its own for the 'rise-fall' pattern and thus predicting more noun phrase responses for this manipulation. Conversely, the 'flat' contour, the 'high flat' contour and the 'fall' contour do not exhibit an own contour on the second part according to

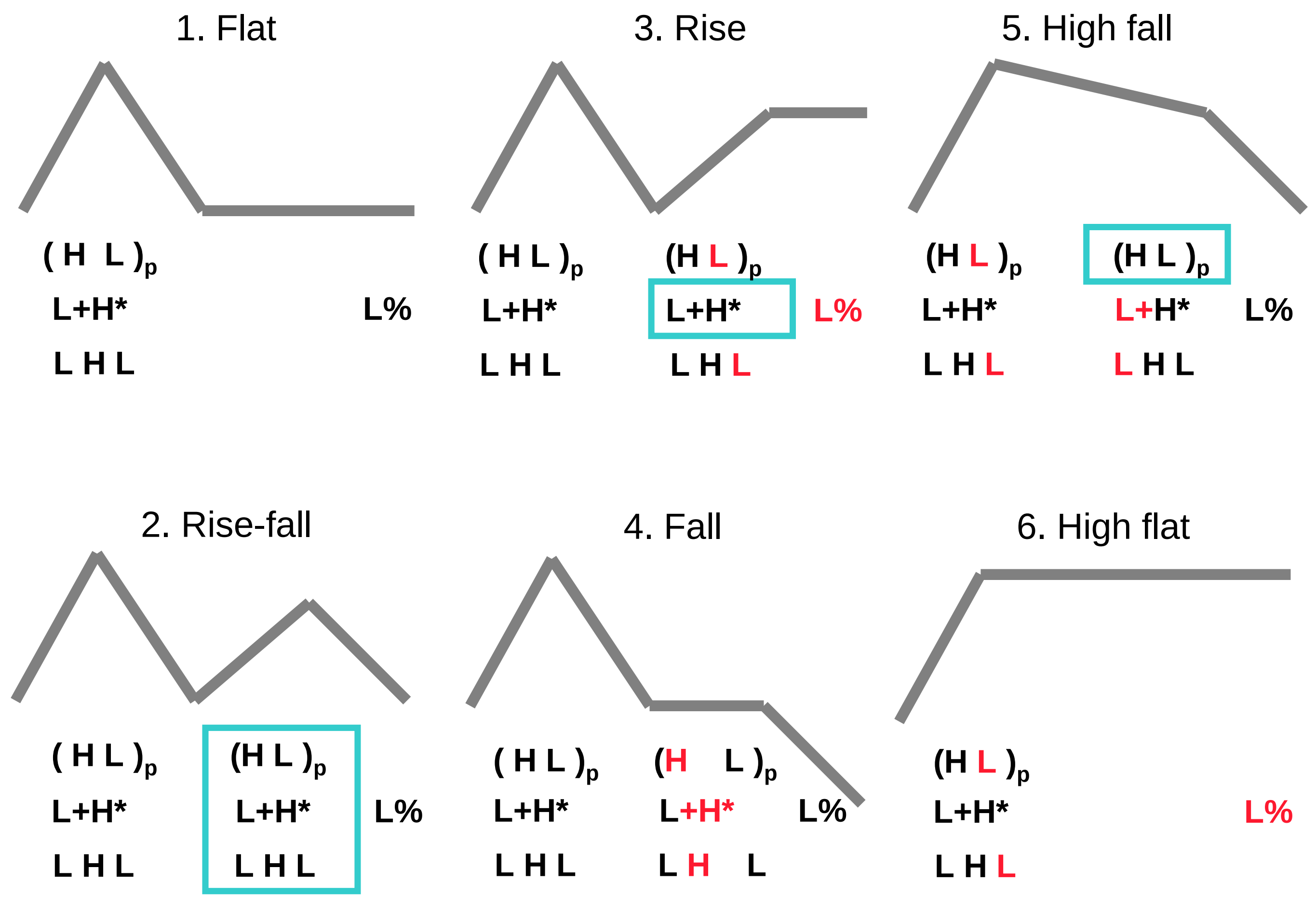

Figure 5.7: Schematic summary predictions based on an analysis in terms of H_pL_p tones (annotations in top row), L+H* accents (middle row) and LHL accents (bottom row).

any of the three analyses and should thus result in a low percentage of noun phrase responses. For 'rise' and 'high fall' contours, the predictions of the three analyses differ.

The analysis in terms of phrase tones interprets the 'rise-fall' contour as the target word constituting an own p-phrase marked by H_pL_p tones. This contour is thus expected to frequently lead to the perception of the compound / noun phrase as consisting of two p-phrases and thus to a noun phrase perception. In contrast, the 'rise' contour does not constitute a complete second p-phrase, since a realisation of the necessary L_p tone is missing. Similarly, the 'fall' contour is analysed as incomplete. Here, the rise-fall contour spanning the preceding p-phrase (the first prosodic word of the compound / noun phrase) is followed by a low plateau and a pitch fall signalling the presence of another L_p tone, but there is no preceding rise and thus no indication for a H_p realisation.[3] The 'high fall' contour, in contrast, is understood as marking the presence of both H_p and L_p on the target word, while the realisation of the H_pL_p tones of the preceding p-phrase is defective, resulting in the lack of an intervening pitch fall. Therefore, listeners are expected to interpret this contour as signalling a noun phrase interpretation significantly more often than the 'flat' pattern, although the incomplete contour of the first part might induce longer response times. Likewise, the contour of the first part of the compound / noun phrase is realised incompletely for the 'high flat' contour, so the effect on response times can be investigated. Since this contour gives no indication for a second p-phrase, the interpretation as a compound should be unambiguous.

Based on the claim by Suomi et al. (2008) that Finnish accents need to be specified as a tritonal LHL instead of just positing targets for either the rise (LH) or the fall (HL), one would assume that only the 'rise-fall' manipulation should induce a significant increase in noun phrase responses. For the 'rise', 'fall' and 'high fall' contours, which are incomplete accents according to this account, responses should more or less be random. The 'high flat' contour signals the presence of one incompletely realised accent on the first part according to this approach.

Following Välimaa-Blum's (1993) description of accents as L+H* (or marginally L*+H), the 'rise-fall' contour should also be the one clearly eliciting noun phrase responses, since participants would expect the accentual rise to be combined with a fall to the phrase-final L% present in her grammar of intonational tones. However, the accent should still be perceivable

[3]An alternative analysis is possible: The final fall could be due exclusively to an L_i boundary tone associated with the i-phrase. To tease the two levels apart, target items would have needed to be embedded in non-final position. However, this would have resulted in very unnatural stimulus sentences. Notice, however, that ascribing the final fall to a higher level boundary tone would mean even more clearly that the compound / noun phrase is realised as only one p-phrase. Thus, it would likewise predict that the 'fall' contour should predominantly induce compound perceptions.

without the boundary tone so that the 'rise' pattern should lead to an equal amount of noun phrase responses as the 'rise-fall' manipulation, although confusion about the incomplete contour might be reflected in longer reaction times. In contrast, the 'fall' and 'high fall' manipulations, missing the rising pitch movement characteristic of an accent, should induce significantly less noun phrase responses following this description of Finnish intonation.

Lastly, an overall bias towards the compound interpretation is expected, since the stimuli were all created from compound recordings, thus retaining temporal and possibly other phonetic cues pointing towards a compound interpretation (recall that stimulus design was purposefully conservative in this respect, as discussed in section 5.2.2).

Also, I suspected that not all items would work equally well as noun phrases. A lower overall amount of noun phrase responses is expected for noun phrases mentioning a possessor not visible on the picture (*amerikan rauta* 'America's iron', *hiiden kirnu* 'the demon's churn', *hullun mylly* 'the crazy person's mill', *juopon nappi* 'the drunkard's button' and *pojan naskali* 'the boy's awl') or in cases where participants would have to infer the identity of the person visible on the picture (*Annan silmä* 'Anna's eye', *anopin kieli* 'mother-in-law's tongue'), although similar cases were included in the practice and filler items to accustom the participants to this. Additionally, segmental assimilation between the two parts of the compound was present in the chosen compound recordings of six items (*hapankaali* 'sauerkraut', *hiidenkirnu* 'pothole', *karhunkieli* 'scouring pad', *kissankello* 'harebell', *kurjenpolvi* 'geranium' and *verenpisara* 'fuchsia'), but only in one of the noun phrase versions (*hapan kaali* 'rotten cabbage'). This probably constitutes a bias towards the compound choice. Also, it was expected that compounds with a higher frequency would be chosen more often.

5.4 Results

5.4.1 Choice

The participants' choices of the pictures showing the referent of the compound or of the noun phrase meaning of the ambiguous sequences are shown in Table 5.2. For all pitch contour patterns, subjects chose the compound interpretation in the majority of trials (3501 cases, i.e. 65% of the responses overall). This is not surprising, since the materials were based on compound recordings and segmental timing had been retained, constituting a bias for the compound choice.

That said, there were significant differences between the conditions. The 'flat' condition with a low, completely flat pitch contour on the target word was the intercept for the best-fitting statistical model (see Table 5.3). In contrast to this condition, participants chose the noun phrase referents more often for almost all other pitch contours. As expected, this effect was significant

Table 5.2: Number and percentage of responses selecting compound vs. noun phrase (NP) referent for the different pitch contour conditions.

	Flat	Rise-fall	Rise	Fall	High fall	High flat
Compound	632(71%)	530(59%)	567(63%)	624(70%)	485(54%)	663(74%)
NP	258(29%)	365(41%)	332(37%)	265(30%)	413(46%)	229(26%)

for the rise-fall contour. The increase in noun phrase choices when only an f_0 rise was present on the target ('rise' contour) was only marginal. However, while the 'rise' pattern only induced marginally more noun phrase choices than the 'flat' pattern (see Table 5.3 again), it also triggered only marginally less noun phrase responses than the 'rise-fall' contour. This emerges from Table 5.4 comparing the non-flat pitch contours with the 'rise-fall' as the intercept.

Table 5.3: Linear mixed-effects model of compound vs. NP choice (random effects: item and trial, pitch condition and logarithmically transformed compound lemma frequency by subject).

Condition	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−1.5662	0.3635	−4.3088	0.0000
Rise-fall	0.7854	0.3343	2.3494	0.0188
Rise	0.5348	0.3083	1.7345	0.0828
Fall	0.0630	0.2003	0.3147	0.7530
High fall	1.1474	0.4425	2.5929	0.0095
High flat	−0.2887	0.1821	−1.5853	0.1129
Trial	0.0000	0.0007	−0.0542	0.9568
log frequency	−0.1812	0.1473	−1.2304	0.2186

A fall from the low pitch level reached at the end of the preceding word alone ('fall' pattern) did not induce a significantly larger proportion of noun phrase choices when compared to the 'flat' pattern (see Table 5.3) and significantly less of them than the 'rise-fall' contour (see Table 5.4).

The pitch contour leading to the highest number of noun phrase choices was the 'high fall' condition, where the fall occurred from a high plateau sustained after the peak of the preceding word. The effect of this fall from a high level was significant compared to the 'flat' condition, and it did not differ significantly from the 'rise-fall' pattern (see Tables 5.3 and 5.4, respectively). The increase in noun phrase choices was even larger if one takes the 'high flat' pattern as a reference (see Table 5.2 and the model of only these two

Table 5.4: Coefficients of the best linear mixed-effects model of compound vs. NP choice, comparing only the four non-flat pitch conditions 'rise-fall' (intercept), 'rise', 'fall' and 'high fall' (random effects: item and pitch condition and frequency by subject).

Condition	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−0.7858	0.3799	−2.0681	0.0386
Rise	−0.2433	0.1481	−1.6434	0.1003
Fall	−0.7078	0.2508	−2.8221	0.0048
High fall	0.3363	0.2211	1.5211	0.1282
log frequency	−0.1856	0.1439	−1.2894	0.1972

Table 5.5: Linear mixed-effects model of compound vs. NP choice, comparing only the two pitch conditions 'high fall' and 'high flat' (random effects: item and pitch condition, trial and logarithmically transformed compound lemma frequency by subject).

Condition	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.0815	0.4511	0.1806	0.8567
High flat	−1.5642	0.4222	−3.7054	0.0002
Trial	−0.0051	0.0027	−1.9217	0.0546
log frequency	−0.1388	0.1301	−1.0667	0.2861

conditions in Table 5.5). The 'high flat' contour, which was included into the experiment specifically to provide a point of comparison for the 'high fall' condition, lead to an even lower number of noun phrase responses than the 'flat' contour, although the difference was not significant.

Regarding the differences between the two parts of the experiment, a comparison between the two flat contours showed more compound responses in the second part of the experiment, whereas the opposite trend can be seen comparing the two falling contours in Table 5.2. Indeed, the factor trial number was not a significant predictor in Table 5.3, indicating no overall trend of participants' choices changing during the course of the experiment. However, an overall trend for a development towards more / less compound choices seemed to occur for some of the participants. The model in Table 5.3, which incorporates the trial number into the random subject effect, was significantly superior to a model without random slopes for trial.

Another factor that had no significant main effect but improved the model when added to the by-subject random factor was frequency. It is likely that participants would have been more inclined to choose the noun phrase referent for items where the compound was very infrequent. A trend into this direction seems visible for some participants in Figure 5.8, which focuses on the items with low compound frequency. In contrast, the three stimuli omitted in the figure induced a high percentage of noun phrase responses for many participants. Overall, the effect of frequency is difficult to assess with the current data set, since the target items were not chosen to represent a balanced selection with respect to this factor.

In spite of this, there were clear by-item differences in the proportion of compound vs. noun phrase choices. This is illustrated in Figure 5.9, which shows that the items naturally split into three groups: Most were interpreted as compounds in the majority of cases across speakers and conditions, some led to an overall similar number noun phrase and compound responses, whereas the smallest group consists of those items that were mostly understood as noun phrases. An example where compound frequency is not a good predictor is the item pair *Joensuu* (a city in Eastern Finland) vs. *joen suu* 'estuary'. Although its compound frequency of over 45 per million is higher than for all other items used in the study, the stimulus was interpreted as a compound far less often then most other items. This could be due to the participants having problems to fast and unequivocally identify a picture of the city centre of Joensuu, which admittedly bears similarities to the city centre of many other middle-sized Finnish towns. Another case in point is the low-frequency compound *pojannaskali* 'cheeky monkey' (0.06 occurrences per million) being picked over its noun phrase equivalent *pojan naskali* 'the boy's awl' in 86% of the stimulus' appearences. Here, possible reasons for disfavouring the noun phrase interpretation could the invisibility of the possessor (i.e. the boy) in the picture or young city dwellers' unfamiliarity with tools like awls. However, compound frequency does seem to

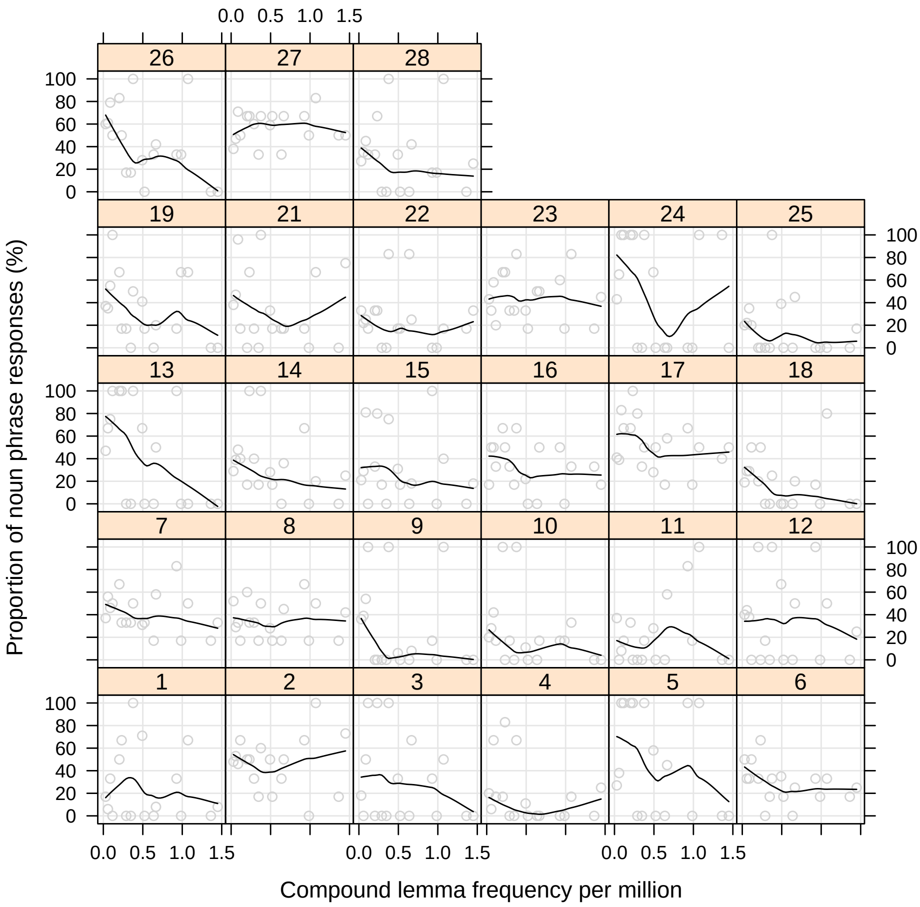

Figure 5.8: Percentage of noun phrase choices in relation to compound frequency by participant. For better display of the majority of the data, the figure does not show data from three items for which the compound frequency was higher than 4 occurrences per million.

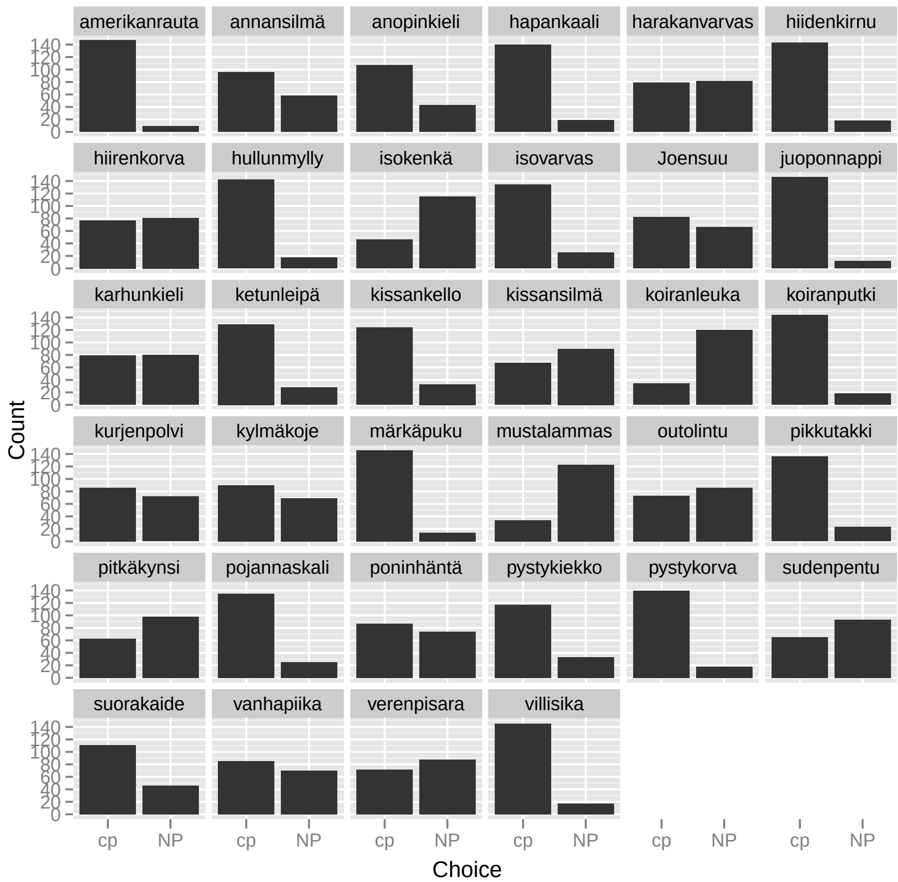

Figure 5.9: Distribution of compound (cp) vs. noun phrase (NP) choices in the whole data set per lexical item.

correlate well with the the general trend towards compound or noun phrase responses for a number of items like the high-frequency compounds *pysty-korva* 'spitz' (vs. 'erect ear') and *pikkutakki* 'blazer, jacket' (vs. 'small coat') or the low-frequency compound *mustalammas* 'black sheep'.

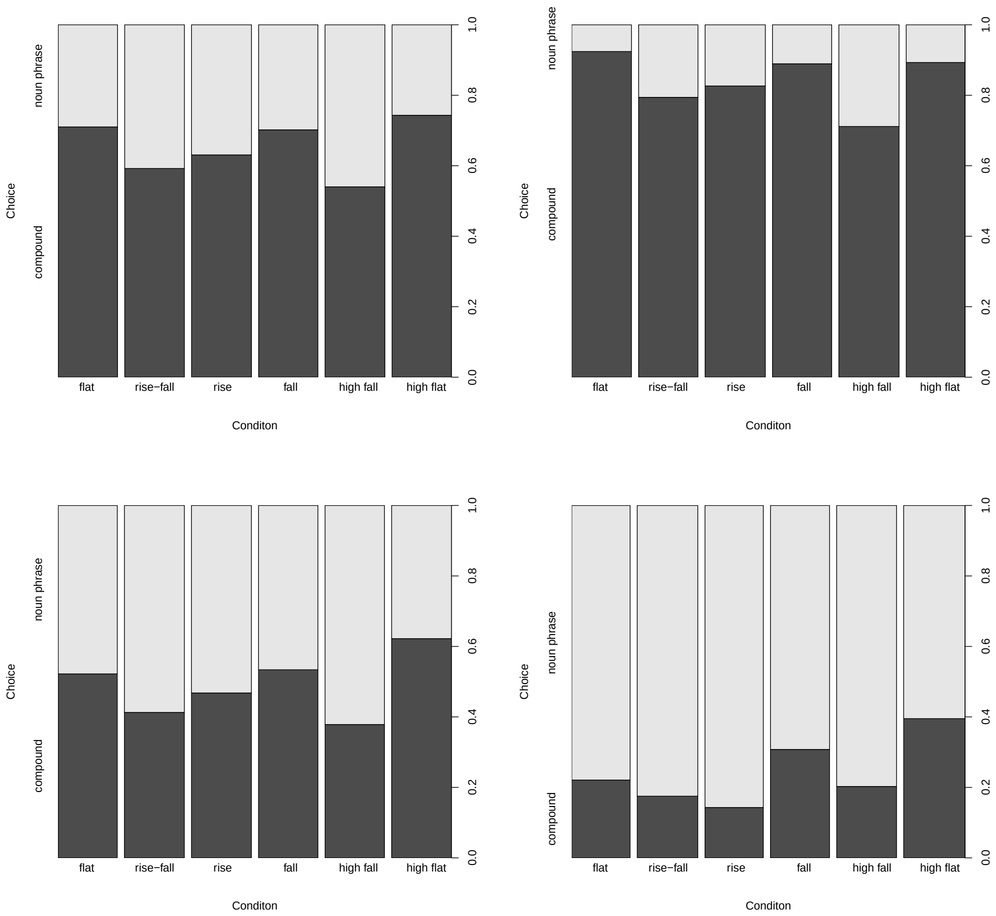

Figure 5.10: Distribution of compound and noun phrase choices for the different pitch contour conditions. Top left panel: All data. Top right panel: Item group C. Bottom left panel: Item group E. Bottom right panel: Item group N.

To assess the differences between the items with a general tendency to be interpreted as compounds or noun phrases, the data was divided into three groups for separate analysis according to item preference, i.e. whether the subjects showed a general preference for one of the two choices per item: Item group C ('compound') consisted of items with overall more than 60% compound responses (18 lexical items, 2838 observations), group E ('equal') contained items that triggered between 60% and 30% compound responses

(13 lexical items, 2053 observations), whereas the three lexical items in group N ('noun phrase') lead to less than 30% compound responses (472 observations). Figure 5.10 compares the effect of the six pitch patterns on the participants' choice of compound or noun phrase referents for these three groups and in the complete data set. It shows that while the general trend towards more or less compound responses differentiated the groups from each other, the relation between the pitch conditions was quite similar.

Table 5.6: Linear mixed-effects model of compound vs. NP choice in item group N, comparing only the two pitch conditions 'rise-fall' and 'rise' (random effects: item and subject).

Condition	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.7886	0.3574	5.0038	0.0000
Rise	0.2521	0.4659	0.5412	0.5884

Table 5.7: Coefficients of the best linear mixed-effects model of compound vs. NP choice in item group N with random effects of item and of pitch condition by subject.

Condition	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.3257	0.6009	3.8706	0.0001
Rise-fall	−0.0480	0.6643	−0.0722	0.9424
Rise	−0.0818	0.7482	−0.1093	0.9130
Fall	−0.2699	0.5563	−0.4852	0.6276
High fall	0.5019	0.7152	0.7017	0.4828
High flat	−1.3727	0.5528	−2.4830	0.0130

The most striking exception was that the 'rise' pattern lead to more noun phrase responses than the 'rise-fall' contour in group N, but this difference was not significant (see the model in Table 5.6). Also, while the 'flat' contour expectedly triggered more compound responses than the rise-falls in all item groups, the difference was clearly smallest and in fact insignificant in group N (see the model in Table 5.7). The number of compound choices for the 'flat' condition being so low in this group, it was even surpassed by that in the 'fall' condition, although this contrast did not reach statistical significance either. Possibly because of the small number of data points, the only condition that differed significantly from the 'flat' intercept was the 'high flat' pattern, triggering more compound responses. Also due to the size of the data set, it was not possible to compute an alternative model with a main or by-subject

Table 5.8: Coefficients of the best linear mixed-effects model of compound vs. NP choice in item group E with random effects of item, pitch condition and frequency by subject.

Condition	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−0.1215	0.2967	−0.4095	0.6822
Rise-fall	0.5812	0.3258	1.7843	0.0744
Rise	0.3088	0.3045	1.0144	0.3104
Fall	−0.0258	0.2337	−0.1104	0.9121
High fall	0.8332	0.3988	2.0894	0.0367
High flat	−0.5176	0.2500	−2.0702	0.0384
log frequency	−0.0219	0.0875	−0.2503	0.8024

Table 5.9: Coefficients of the best linear mixed-effects model of compound vs. NP choice in item group C with random effects of item and of pitch condition and frequency by subject.

Condition	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−3.1187	0.3066	−10.1721	0.0000
Rise-fall	0.8566	0.3988	2.1479	0.0317
Rise	0.6381	0.3575	1.7852	0.0742
Fall	0.0644	0.3337	0.1930	0.8469
High fall	1.4639	0.4901	2.9870	0.0028
High flat	0.3304	0.2705	1.2215	0.2219
log frequency	−0.1562	0.1154	−1.3539	0.1758

effect of trial number. Adding a term for compound frequency did not yield an improvement.

For item preference group E, the random effects were more complex again (see Table 5.8). The picture emerging from this item group was similar to the result of analysing the complete data set. However, the difference between the 'flat' and the 'rise-fall' condition was only marginally significant, whereas the higher amount of compound responses to the 'high flat' stimuli did reach significance when only group E was considered.

Non-surprisingly, the analysis of the largest item group C showed the same significant effects as the data set on a whole (see Table 5.9). Interestingly, also the 'high flat' condition had an—insignificantly—higher number of noun phrase choices in group C, inverting the effect found for the data set as a whole.

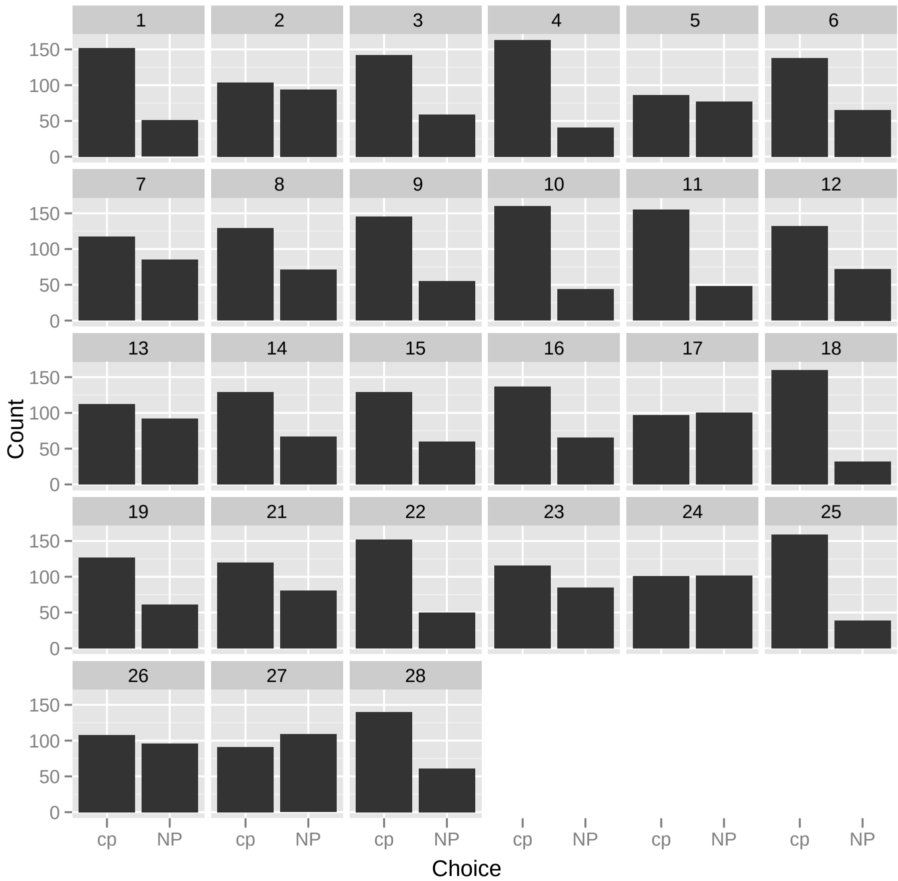

Figure 5.11: Distribution of compound (cp) vs. noun phrase (NP) choices in the whole data set per subject.

A similar division into groups can be made for the subjects. Figure 5.11 shows that while participants 17, 24 and 27 chose more noun phrase than compound referents on a whole, the difference was small and none of the participants exhibited a strong bias towards noun phrase responses. Thus, no subset corresponding to item group N could be defined. Instead, subjects were divided into two subsets, corresponding to item preference groups C and E above, according to whether or not they had chosen compound referents in a large majority of their responses. Again, 60% was used as a dividing line.

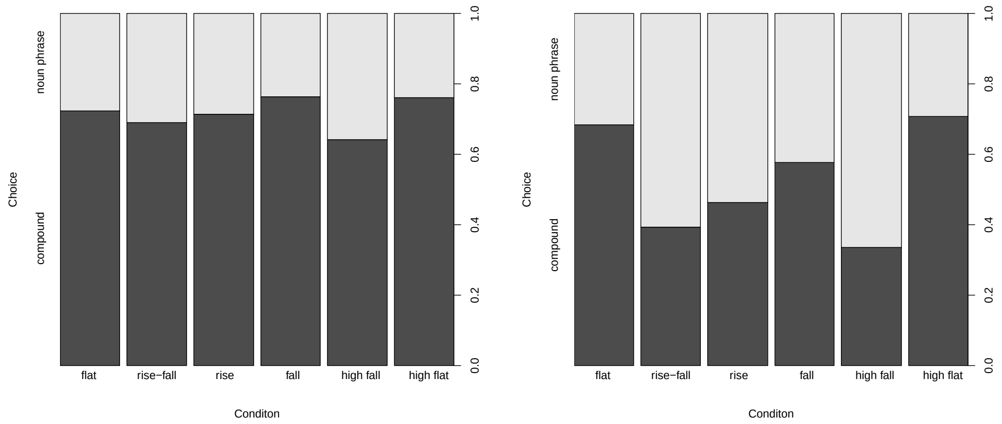

Figure 5.12: Distribution of compound and noun phrase choices for the different pitch contour conditions for subject group C (left panel) and E (right panel).

The effect of pitch condition for the two subject preference groups is compared in Figure 5.12. Both plots look similar to those in Figure 5.10. However, the differences between the conditions are very small in the left panel. For subject preference group C with more than 60% compound responses (18 subjects, 3591 observations), no significant main effects were found (see Table 5.10).

The differences between the pitch patterns were much clearer for the second group of subjects (group E: 9 subjects, 1772 observations). These participants chose more noun phrase referents on a whole, but the right panel of Figure 5.12 shows that this increase was not equally distributed over the pitch conditions. The effect only occurred in conditions with pitch movements on the target word, while the two flat patterns did not differ from each other. Significantly more noun phrase choices than in the 'flat' condition arose for the 'rise-fall', 'rise', 'fall' and 'high fall' patterns (see Table 5.11). This means that more of the effects that were already visible for the complete data set reached significance when disregarding subjects with

Table 5.10: Coefficients of the best linear mixed-effects model of compound vs. NP choice in subject group C with random effects of item and of pitch condition and frequency by subject.

Condition	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−1.6663	0.3785	−4.4020	0.0000
Rise-fall	0.2378	0.2823	0.8425	0.3995
Rise	0.1020	0.2344	0.4351	0.6635
Fall	−0.2447	0.2354	−1.0394	0.2986
High fall	0.5300	0.4465	1.1870	0.2352
High flat	−0.2831	0.2335	−1.2125	0.2253
log frequency	−0.1805	0.1548	−1.1664	0.2434

Table 5.11: Coefficients of the best linear mixed-effects model of compound vs. NP choice in subject group E with random effects of item and pitch condition and frequency by subject.

Condition	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−1.4710	0.4657	−3.1585	0.0016
RiseFall	1.8968	0.7262	2.6119	0.0090
Rise	1.4431	0.7342	1.9655	0.0494
Fall	0.7140	0.3078	2.3200	0.0203
High fall	2.4044	0.8724	2.7561	0.0059
High flat	−0.2788	0.2979	−0.9359	0.3493
log frequency	−0.2053	0.1620	−1.2670	0.2052

a strong tendency towards compound choices across all conditions. As in the complete data set, the amount of noun phrase choices was only marginally smaller for the 'rise' contour when compared to the 'rise-fall' pattern, but significantly smaller for the 'fall' contour, while the 'high fall' induced insignificantly more noun phrase responses (see Table 5.12).

Figure 5.13 confirms that some subjects were more sensitive to the experimental conditions, while others clearly adopted an overall strategy to choose compound referents—possibly additionally influenced by further factors like compound frequency and trial number. Three of the participants—05, 13 and 24—obviously responded randomly. They chose the noun phrase referent for 50-43%, 50-44% and 52-50% of the trials in the different pitch conditions, respectively.[4]

[4]Another model was fit to the data from subject group E after removing the randomly responding participants 05, 13 and 24. While the effects became stronger overall, as

Table 5.12: Coefficients of the best linear mixed-effects model of compound vs. NP choice in subject group E, comparing only the four non-flat pitch conditions ‘rise-fall’ (intercept), ‘rise’, ‘fall’ and ‘high fall’ (random effects: item and pitch condition and frequency by subject).

Condition	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.4213	0.5114	0.8239	0.4100
Rise	−0.4436	0.2688	−1.6505	0.0988
Fall	−1.1563	0.5787	−1.9981	0.0457
High fall	0.4853	0.3150	1.5405	0.1234
log frequency	−0.1986	0.1608	−1.2352	0.2167

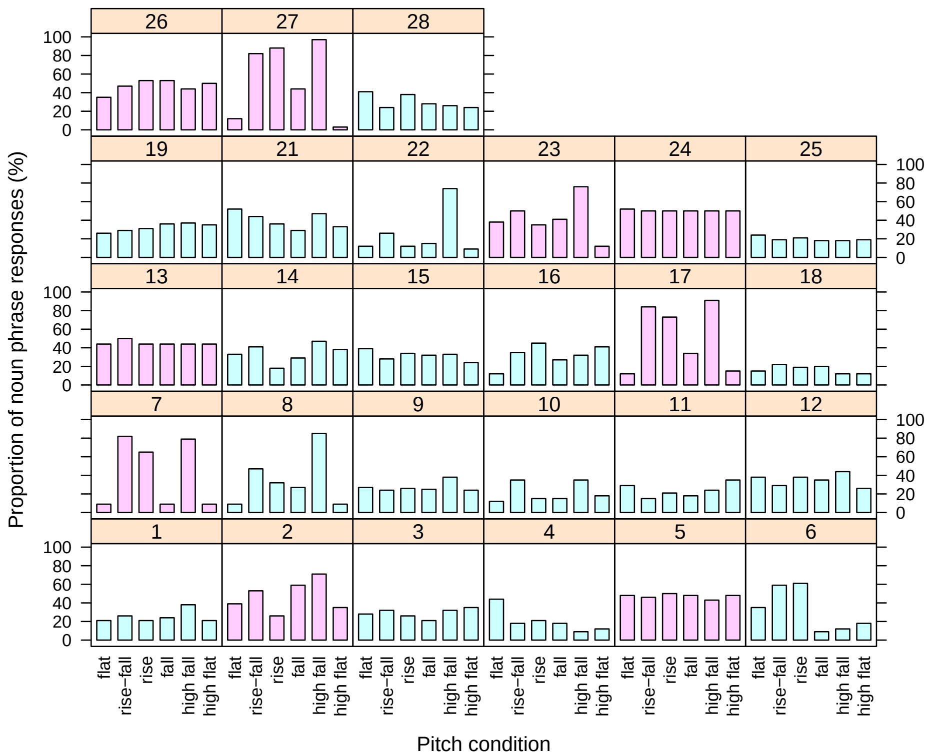

Figure 5.13: Percentage of noun phrase choices in different pitch conditions per subject, plotted with turquoise bars for subjects from subject preference group C and with pink bars for subjects from group E.

On a whole, the figure shows that the division into two groups along the line of 60% overall compound choices was quite efficient in separating subjects who used a strategy from those who were more sensitive to the experimentally induced variations. Although for example subject 08 who did react to the pitch manipulations is unnecessarily excluded, this principled and conservative selection was preferred over hand-picking 'the good subjects'. This decision is supported by the fact that all linguistically interesting experimental conditions differed for subject group E (see Table 5.11 again), whereas group C did not exhibit any significant differences (see Table 5.10).

Also in line with the statistical modelling discussed above, Figure 5.14 shows that items groups C, E and N differed with respect to the overall tendency for or against noun phrase choices, but not with respect to their sensitivity towards the pitch conditions.

5.4.2 Reaction times

How fast the participants responded was influenced directly by the pitch contour of the stimulus: The average reaction time was significantly longer in the 'flat' condition than for all other pitch contours. Moreover, the response times significantly differed between trials where the participants chose compound referents and those where they chose noun phrase referents, with noun phrase choices generally taking longer (see Tables 5.13 and 5.14). This could be interpreted as an additional indirect effect of the pitch contour if one argued that noun phrase choices, which were more often triggered by certain pitch contours, took longer to execute. The opposite conclusion is also possible, i.e. in contrast to assuming that the choice affects the reaction time, the reaction time could also affect the choice. Thus, it could be that the longer it took the participant to respond, the more likely it was that she or he would choose the noun phrase referent.

However, no significant interaction between the pitch contour of the stimulus sound and the participants' choice was found to affect the reaction times: The model in Table 5.14 was not significantly improved by including such an interaction. This is sightly surprising considering Figure 5.15 showing that the differences between reaction times of compound and noun phrase choices

expected when removing noise from the data, and no complex random effects structure was necessary (by subject random slopes for frequency or trial) or computable (by-subject random slopes for pitch condition), all significant effects were the same as in the previous model in Table 5.11:

Condition	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	−1.3228	0.2412	−5.4852	0.0000
Rise-fall	2.1036	0.2393	8.7917	0.0000
Rise	1.6228	0.2327	6.9734	0.0000
Fall	0.8488	0.2328	3.6455	0.0003
High fall	2.6534	0.2513	10.5595	0.0000
High flat	−0.2407	0.2536	−0.9492	0.3425

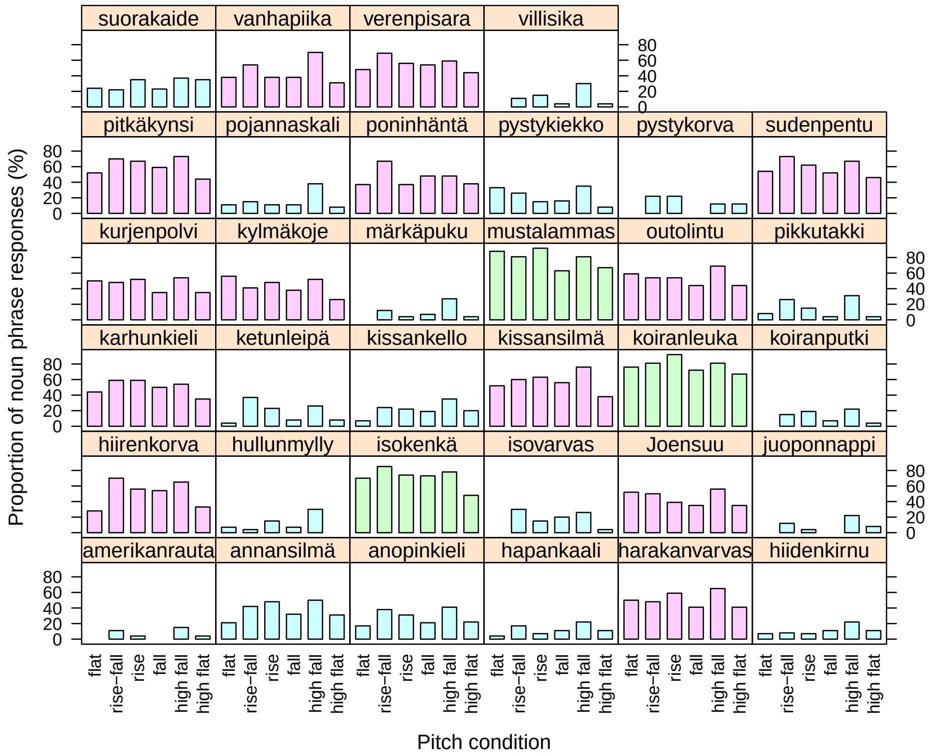

Figure 5.14: Percentage of noun phrase choices in different pitch conditions per item, plotted with turquoise bars for items from item preference group C, pink bars for items from group E with light green bars for items from group N.

Table 5.13: Mean and standard deviation (SD) of reaction time (in ms) for compound and noun phrase responses in different pitch conditions.

	Compound choices		Noun phrase choices	
Condition	*Mean*	*SD*	*Mean*	*SD*
Flat	1102	476	1143	509
Rise-fall	828	401	1035	504
Rise	877	441	1046	481
Fall	887	455	1001	503
High fall	804	416	1050	471
High flat	877	443	893	455

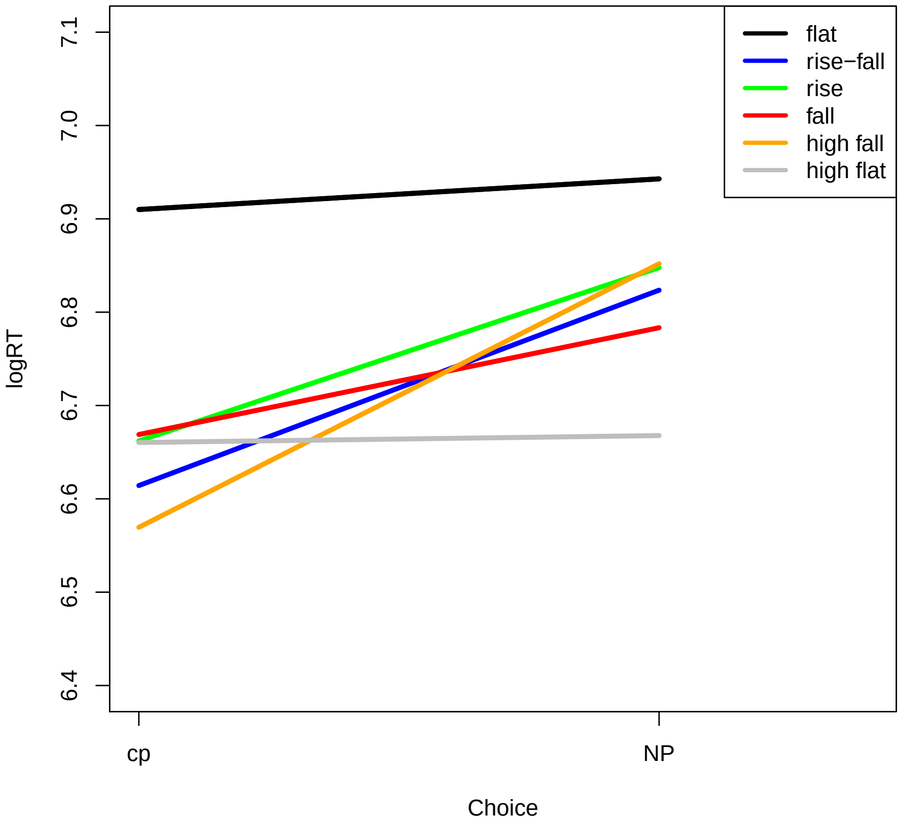

Figure 5.15: Interaction plot: Mean logarithmically transformed reaction times for compound (cp) and noun phrase (NP) responses to different pitch contours.

were consistently larger for the non-flat conditions. Interactions were found with the grouping as C, E or N according to overall preference for noun phrase or compound choices per item and per subject (see section 5.4.1; three-way or even four-way interaction models were not significantly better). The overall effect causing longer reaction times for noun phrase choices was less strong for subjects who had no clear general preference for either compounds or noun phrase responses (subject preference E 'equal'). The same was true of items which induced both responses about equally often (item preference group E 'equal'). For items which elicited more noun phrase choices overall (item group N), these responses were even faster than the compound choices. The common denominator of these interactions seems to be that more frequent responses took less time.

By-item preference also had a main effect: It took the participants significantly longer to decide in case of items for which they had no clear preference and also for those where they chose the noun phrase response most often. This can be explained by the overall tendency towards compound choices, which made deviating choices more time-intensive. Likewise, subjects who showed no strategy of mostly choosing compounds exhibited longer overall reaction times, but this effect did not reach significance. Also, of the interactions between pitch contours and item preference groupings only one—an increase of the general quickening effect of 'rise-fall' contours for item group N—was marginally significant, while adding a pitch contour by subject preference interaction only marginally improved the fit and this term was thus not included in the model.

The last main factor included in the model in Table 5.14, trial number, did not have a main effect. Thus, no general trend towards longer or shorter reaction times over the course of a session could be identified. However, tendencies can clearly be seen for some of the subjects in Figure 5.16, e.g. a familiarisation for participants 18 and 22, while a slight effect of fatigue is visible for participants 4 and 24. Consequently, including by-subject random slopes for trial significantly improved the model.

Another by-subject random effect was also included, namely that of pitch condition. This means that main effects of this factor affected all participants slightly differently. The most clearly visible difference in Figure 5.17 is the much larger reaction time to the 'flat' contour for some subjects. As the reader can verify in comparison with Figure 5.13, these participants (I identified 03, 11, 12, 14, 16, 18, 19, 21, 24, 25 and 28) were all part of subject group C, with the only exception of subject 24. Thus, subjects who mostly selected compound interpretations were particularly slow in reacting to the 'flat' contour.

Therefore, a separate model was fit to the data from subject group E, i.e. the participants who were more sensitive towards the experimental conditions (see section 5.4.1). The model did not display the same differences between the pitch conditions as the model of the complete data set, although a trend

Table 5.14: Coefficients of the best linear mixed-effects model of reaction times for the complete data set, with random effects of item and of pitch condition and trial by subject.

Condition	Estimate	Std. Error	t value
(Intercept)	6.8902	0.0692	99.5143
Rise-fall	−0.2356	0.0537	−4.3905
Rise	−0.2271	0.0436	−5.2144
Fall	−0.2211	0.0497	−4.4512
High fall	−0.2580	0.0595	−4.3336
High flat	−0.3068	0.0582	−5.2680
Choice 'noun phrase'	0.2457	0.0220	11.1800
Item group E	0.1265	0.0387	3.2678
Item group N	0.2730	0.0720	3.7913
Subject group E	0.1061	0.0992	1.0699
Trial	−0.0005	0.0003	−1.5354
Rise-fall : Item group E	−0.0014	0.0354	−0.0408
Rise : Item group E	0.0503	0.0353	1.4256
Fall : Item group E	−0.0110	0.0354	−0.3106
High fall : Item group E	−0.0521	0.0356	−1.4656
High flat : Item group E	−0.0369	0.0354	−1.0435
Rise-fall : Item group N	−0.1083	0.0608	−1.7820
Rise : Item group N	0.0490	0.0612	0.8006
Fall : Item group N	−0.0916	0.0610	−1.5021
High fall : Item group N	−0.0789	0.0609	−1.2958
High flat : Item group N	−0.0644	0.0608	−1.0593
Choice 'noun phrase' : Subject group E	−0.1029	0.0224	−4.5851
Item group E : Choice 'noun phrase'	−0.1914	0.0253	−7.5736
Item group N : Choice 'noun phrase'	−0.3435	0.0442	−7.7641

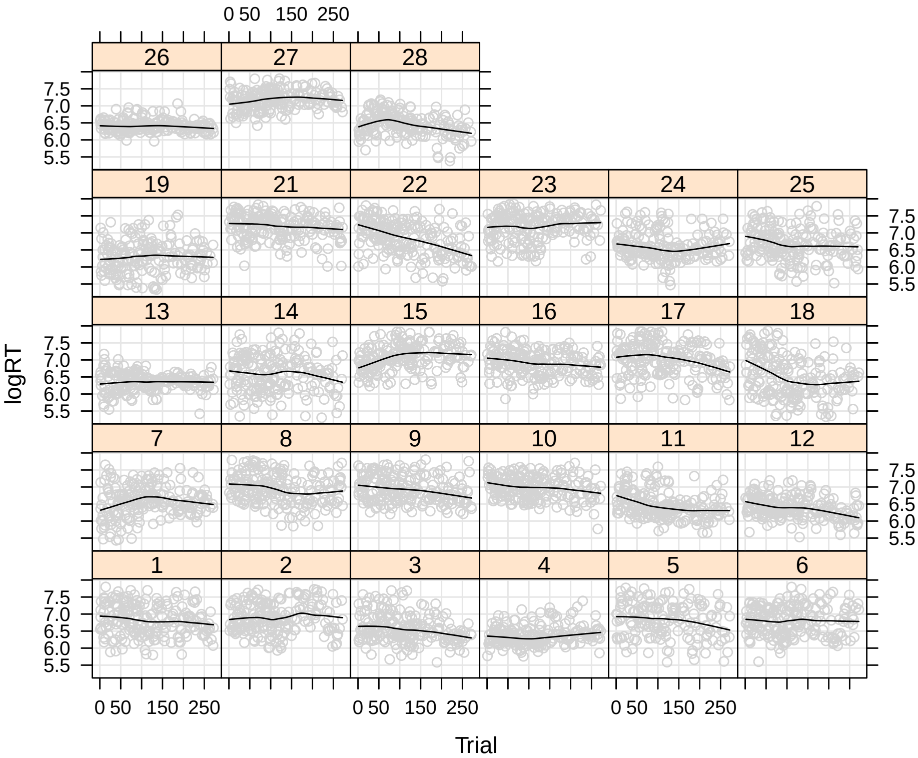

Figure 5.16: By-subject effect of trial on logarithmically transformed reaction times.

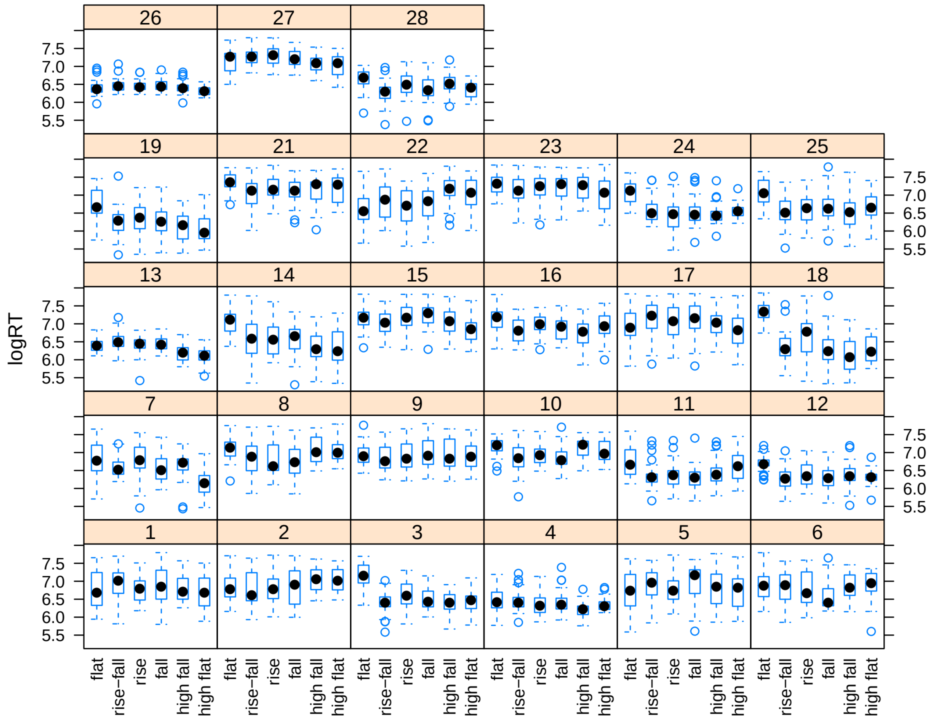

Figure 5.17: Boxplot showing logarithmically transformed reaction times to different pitch contours per subject.

Table 5.15: Coefficients of the best linear mixed-effects model of reaction times for subject preference group E, with random effects of item and of pitch condition and trial by subject (N = 1772).

Condition	Estimate	Std. Error	t value
(Intercept)	6.9325	0.1040	66.6273
Rise-fall	−0.1884	0.1075	−1.7528
Rise	−0.1595	0.0845	−1.8865
Fall	−0.1360	0.0908	−1.4981
High fall	−0.1940	0.0786	−2.4678
High flat	−0.3250	0.0875	−3.7129
Choice ‘noun phrase’	0.0964	0.0520	1.8545
Item group E	0.0950	0.0379	2.5032
Item group N	0.1966	0.0743	2.6470
Trial	−0.0003	0.0006	−0.4738
Rise-fall : Choice ‘noun phrase’	0.0776	0.0624	1.2442
Rise : Choice ‘noun phrase’	0.1010	0.0616	1.6394
Fall : Choice ‘noun phrase’	0.0241	0.0617	0.3901
High fall : Choice ‘noun phrase’	0.0126	0.0641	0.1971
High flat : Choice ‘noun phrase’	0.0302	0.0659	0.4590
Choice ‘noun phrase’ : Item group E	−0.1624	0.0404	−4.0197
Choice ‘noun phrase’ : Item group N	−0.3580	0.0720	−4.9708

for longer response times to the 'flat' pattern was still visible (see Table 5.15). However, only the two contours that appeared exclusively in the second part of the experimental sessions, 'high fall' and 'high flat', elicited significantly faster reaction times, so that this could possibly be explained as an effect of familiarisation, although the same difference was marginally significant for the 'rise' and perhaps for the 'rise-fall' contour.

Also, reaction times for noun phrase choices were only marginally larger in this subset of the data. As in the model for the complete data set, this effect was affected by an interaction with the item group: For items in groups E and N, which were more often categorised as noun phrases, the noun phrase choices were also faster. Interactions between choice and pitch contour were all insignificant but a version of the model without this term failed to converge, so that it was impossible to verify that the interaction term was superfluous.

Lastly, a model was fit to see whether participants were faster in categorising the target as a noun phrase when it carried a 'rise-fall' contour than for the other non-flat patterns. However, no significant differences were found between the pitch conditions—'rise-fall', 'rise', 'fall' and 'high fall'—in the relevant subset of the data (see Table 5.16). In fact, the only significant effect on the reaction times for noun phrase responses to these pitch conditions was that participants where faster in making this choice for items that they generally categorised as noun phrases more often, i.e. item groups E and N. This effect was less strong for subjects who had a weaker overall tendency towards compound choices, i.e. subject group E.

Table 5.16: Coefficients of the best linear mixed-effects model of reaction times for noun phrase responses to non-flat contours (all subjects and items), with random effects of item and of pitch condition and trial by subject.

Condition	Estimate	Std. Error	t value
(Intercept)	6.9516	0.1122	61.9615
Rise	0.0793	0.0451	1.7574
Fall	0.0005	0.0336	0.0161
High fall	−0.0389	0.0628	−0.6186
Item group E	−0.1462	0.0408	−3.5839
Item group N	−0.1718	0.0536	−3.2025
Subject group E	−0.0484	0.1333	−0.3631
Trial	−0.0004	0.0004	−1.0527
Item group E : Subject group E	0.1110	0.0476	2.3302
Item group N : Subject group E	0.0210	0.0621	0.3383

5.5 Summary and discussion

This experiment investigated whether six different manipulated pitch contours on the second part of a compound could elicit a perception of a second p-phrase and thus a categorisation of the complete string as a noun phrase. In spite of an overall bias towards compound responses and difference between subjects and items, a seizable amount of noun phrase interpretations appeared in the responses to some of the pitch conditions. A first finding is therefore that it was possible to 'create' perceptions of noun phrases on the basis of recordings of their compound equivalents by f_0 manipulations without further adjustments. Crucially, the experimental design was successful in assessing differences between the six contours—the patterns 'flat', 'rise-fall', 'rise', 'fall', 'high fall' and 'high flat'. Compared to the flat pitch naturally occurring on the second part of compounds, the 'rise-fall' pitch pattern naturally occurring for noun phrases, as well as the 'high fall' pattern elicited significantly more noun phrase responses (see the overview in Table 5.17). By contrast, the 'high flat' and 'fall' contours did not differ significantly from the 'flat' pattern, but resulted in significantly less noun phrase responses than the 'rise-fall' contour. The 'rise' contour appears between these two groups in Table 5.17, as it did not trigger significantly more noun phrase choices than the 'flat' contour, but also not significantly less of them than the 'rise-fall' contour.

Table 5.17 further compares these results to the predictions from three accounts of Finnish intonation as laid out in section 5.3: first, the analysis in terms of H_pL_p phrase tones advocated here, second, Välimaa-Blum's (1988; 1993) inventory containing L+H* and L*+H accents and third, Suomi et al.'s (2008) description of LHL accents. As detailed in section 5.3, each of the three approaches predicted that some of the pitch patterns would induce more noun phrase responses and others less. The contrast between more noun phrase choices for the 'rise-fall' pattern and less noun phrase choices for target words with completely flat pitch was expected on the basis of all three accounts. With respect to the task at hand, it did not make a difference whether the pitch contour of the relevant word was flat on a low or on a high level.

The analysis in terms of H_pL_p phrase tones hypothesised that only the 'rise-fall' and the 'high fall' pattern should lead to substantially more noun phrase choices, while the 'rise' and 'fall' contours should cluster with the 'flat' and 'high flat' pattern in inducing mostly compound responses. These hypotheses were largely in agreement with the results, as Table 5.17 illustrates: The 'rise-fall' and 'high fall' pitch patterns did indeed elicit significantly more noun phrase responses than the 'flat' contour, while the 'rise', 'fall' and 'high flat' manipulations did not. Only the 'rise' contour did not fit this predicted binary classification quite as neatly as the other manipulations.

Table 5.17: Summary of results regarding choice of compound or noun phrase interpretations for the data set on a whole compared to the predictions based on three different accounts of Finnish intonation (cf. Table 5.1).

	Predictions			**Results**
	$\mathbf{H_pL_p}$	**L+H* / L*+H**	**LHL**	
More NP responses ↑	Rise-fall	Rise-fall	Rise-falls	High fall
	High fall	Rise		Rise-fall
			Rise	Rise
	Rise	Fall	Fall	
	Fall	High fall	High fall	Fall
	High flat	High flat	High flat	Flat
Less NP responses ↓	Flat	Flat	Flat	High flat

The analysis in terms of rising L+H* or L*+H accent predicted that only the 'rise-fall' and the 'rise' pattern should increase the proportion of noun phrase choices compared to the 'flat' contour, while the 'fall', 'high fall' and 'high flat' pattern should not. The results did not match these predictions very well. Crucially, the 'rise' pattern did not induce an increase in noun phrase categorisations. This finding does not fit the hypothesis that the rising part of Finnish rising-falling pitch contours is a tonally specified accent, whereas the fall is just a transition to the leading tone of the next accent or boundary tone, as suggested by Välimaa-Blum (1993). Moreover, the high proportion of noun phrase responses induced by the 'high fall' pattern—exceeding that caused by the 'rise-fall' contour—is difficult to reconcile with this approach. The 'high fall' manipulation only realised a pitch fall from a high level on the target word (the second prosodic word of the ambiguous compound / noun phrase stretch). This fall was preceded by high plateau and, before that, a rise at the beginning of the first prosodic word of the compound / noun phrase. In terms of L+H* accents, the first part would thus carry a completely realised accent, while the second part would not, as annotated in Figure 5.7. Thus, it should have lead to compound identifications just as unambiguously as the 'flat' contour, although possibly some uncertainty might have been induced by the high plateau. Instead, this contour, lacking the—under the L+H* / L*+H crucial—pitch rise on the target word, induced the highest proportion of noun phrase categorisations.

According to Suomi et al.'s (2008) assertion that Finnish accents are uniformly of the type LHL, it was predicted that only the 'rise-fall' pattern would lead to a higher proportion of noun phrase responses. This hypothesis was not born out. Instead, one would be lead to conclude that also incomplete accent realisations were sufficient for participants to detect the presence of an accent and accordingly choose the picture of the noun phrase referent in the present study. The patterns ‘rise’ and ‘fall’, which would be seen as ‘incomplete accents’, induced less noun phrase choices than the ‘complete rise-fall’. However, this difference was only marginal for the ‘rise’ contour. Crucially, the ‘high fall’ contour also lacking the pitch rise did not differ significantly from the ‘rise-fall’ contour and triggered even more noun phrase choices than the supposedly complete accent realisation.

In sum, regarding the question of tonal specification, the analysis in terms of H_pL_p seemed to best account for the data, since the pitch patterns eliciting a significant increase of noun phrase responses were those realising a high and low tone on the target word, whereas the presence or absence of a preceding low tone was not crucial. Additionally, the results of the present experiment are also relevant to the other research questions of this thesis as stated in (10).

The results for reaction times can mainly be summed up by saying that more frequent choices required less time. For instance, noun phrase responses, being overall less frequent than compound responses, coincided with longer response times, but this effect was weaker for subjects who had a weaker overall tendency towards compound choices. In this light, it is not clear why reaction times were longest for the ‘flat’ contour, which predominantly elicited compound responses. One potentially interesting finding is that subjects were equally fast in choosing the noun phrase referent in response to the ‘rise’, ‘fall’ or ’high fall’ pattern as they were when making the same response to a ‘rise-fall’ stimulus. It is possible that the method employed here was simply not sensitive enough to capture existing differences. However, if the finding were to be trusted, this would mean that participants had no trouble categorising these incomplete contours as realisations of noun phrases. For the ‘rise’ and ‘fall’ contours, the number of noun phrase categorisations was not very high, but at 332 and 265 cases not exactly low, either (recall Table 5.2). For the ‘high fall’ manipulation, the proportion of relatively fast noun phrase responses approached 50%.

Thus, participants reacted to incomplete ‘rise-fall’ contours and frequently chose noun phrase interpretations in response. Moreover, it is worth noting that most of the manipulations resulted in contours that do not feature in previous descriptions in the literature and might not naturally occur in Finnish. However, several native speakers have told me that none of the pitch patterns used in this experiment sounded weird or distinctly unnatural and that even ranking their (un)naturalness was difficult. Furthermore, when I asked whether some of these contours implied different meanings, I was told that the ‘high flat’ contour carried an implication of continuation, but that the other contours had no connotations of this kind.[5] In this regard, the findings for subject group E, the participants more sensitive to the experimental manipulations, were especially interesting. For these subjects, the increase in noun phrase choices was significant for all four types of non-flat pitch contours (i.e. for the patterns 'rise-fall', 'rise', 'fall' and 'high fall'). In the complete data set including subjects who adopted a strong overall strategy to choose the compound response, not all of these effects reached significance.

That participants reacted to a variety of incomplete tonal contours indicates a difference from the perception of accent in intonation languages. For example speakers of a Germanic language generally either perceive an accent being realised on a certain word or they do not perceive this accent being present on this word—this kind of categorical distinction has for example been exploited in perception experiments rather similar to the current one, which have been used as evidence pertaining to the on-ramp vs. off-ramp debate (see e.g. Gussenhoven, 2008; Chen, 2011). In contrast, the participants of the present study were more likely to categorise the ambiguous stretch as a noun phrase when any kind of pitch movement was present on the second prosodic word. This kind of variability is more in line with a phrase language account as suggested here than with an analysis in terms of accents. Whereas no all-or-none effect in line with an accent interpretation could be found, the results are consistent with the following interpretation: When participants perceived more pitch movements than expected for the realisation of one p-phrase, they were more likely to attribute the superfluous movements to the existence of a second p-phrase and thus to choose the noun phrase interpretation. This agrees with the variability regularly found in phrase languages (see Féry, 2010, on variation in phrasing and tonal realisation in Indian phrase languages).

Finally, a few comments on referring to the p-phrase as the prosodic level distinguishing compounds and noun phrases is in order. Compounds have been analysed employing a recursive prosodic word structure in many languages, including Finnish (Karvonen, 2005, see). The prosodic difference between such a prosodic word (PW) containing two PWs and a noun phrase containing the same two words is illustrated for a minimal pair in (28). By contrast, the present suggestion is that at least short two-part compounds consist of two prosodic words forming one p-phrase, as illustrated in (29a), while corresponding noun phrases form two separate p-phrases, as shown in (29b).[6] As discussed in Chapter 1, this thesis assumes that prosodic words are the domain of syllabification, vowel harmony and other segmental phenomena, while p-phrases are most prominently marked by the tonal contours

[5] In addition to my colleagues in Helsinki, I thank attendees of the *Nordic Prosody XI* conference in Tartu, August 2012, for sharing their intuitions with me. Naturally, I assume full responsibility for my interpretation.

[6] The materials of the present

which were the focus of the present chapter, analysed as realisations of H_pL_p tones here. Since compounds do not constitute a single domain of vowel harmony (notice the occurrence of front-harmonic *ä* and back-harmonic *u* in the example below), the analysis as separate PWs seems straightforward (but see footnote 4). More importantly, the tonal movements analysed as H_pL_p tones here can span several morpho-syntactic words and even whole sentences, more in line with the ascription to the phonological phrase level suggested here than with attributing them to the PW level. Interestingly, if the compound / noun phrase distinction lies in the distribution of p-phrases, Niemi's (1984) findings regarding the phonetic cues to the distinction imply that there are further correlates of p-phrases than H_pL_p tones, especially duration and intensity.

(28) a. ((märkä)$_{\mathrm{PW}}$(puku)$_{\mathrm{PW}}$)$_{\mathrm{PW}}$
märkä puku
wet suit
'wetsuit' (garment worn for diving)

b. ((märkä)$_{\mathrm{PW}}$ (puku)$_{\mathrm{PW}}$)$_{\mathrm{p}}$
'wet suit' (a suit that got wet)

(29) a. ((märkä)$_{\mathrm{PW}}$(puku)$_{\mathrm{PW}}$)$_{\mathrm{p}}$
'wetsuit' (garment worn for diving)

b. ((märkä)$_{\mathrm{PW}}$)$_{\mathrm{p}}$ ((puku)$_{\mathrm{PW}}$)$_{\mathrm{p}}$
'wet suit' (a suit that got wet)

In conclusion, the materials presented in this chapters support an analysis of Finnish rise-fall pitch contours as realisations of H_pL_p tones associated with p-phrases. They further show substantial differences between Finnish and prototypical intonation languages like English, Dutch or German, in line with the phrase language account suggested in this thesis.

Chapter 6

Conclusion

6.1 Summary

This thesis has presented studies on phrasing and intonation in Finnish, intended to address the research questions below.

(30) 1. Is Finnish phrase-level intonation indeed largely uniform or are there meaningful phonological contrasts that warrant the assumption of contrasting accents?

2. What is the shape and tonal specification of the pitch contour(s)?

3. What is the position of Finnish in a prosodic typology?

After the first chapter provided an overview of the thesis as a whole and an introduction to the relevant questions and the suggested prosodic analysis, Chapter 2 presented Production Experiment I, designed to address the first two research questions in (30). The study investigated prosodic effects of variations in focus condition, position / grammatical function and first syllable vowel quantity. It found effects of all three factors and modelled them in terms of p-phrases marked by $\mathrm{H_pL_p}$ tones and i-phrases right-aligned with focus constituents.

Chapter 3 addressed a possible alternative interpretation of the findings in Chapter 2, namely postulating a distinction between nuclear H*L accents and non-nuclear LH*. It refuted this alternative through a more detailed analysis of peak alignment in the data from Production Experiment II. Additionally, the chapter presented a pilot study finding an effect of peak height, but not of peak timing for the perception of constituents as narrowly focused. Finally, a further analysis of the data from Chapter 2 revealed no effect of stress on intensity.

The second production experiment, discussed in Chapter 4, included items with a wider range of different syllable structures than Production Experiment I. It explained differences between these syllable structures as generally due to durational effects of the number and phonological quantity of segments. It also reaffirmed the suitability of an analysis in terms of phrasing and specified the association of the phrase tones H_pL_p as associated with the beginning and the end of the p-phrase, respectively.

The main purpose of the processing study in Chapter 5 was to assess the second research question by testing which tonal movements are necessary and sufficient for native speakers of Finnish to perceive a word as a p-phrase of its own. The underlying hypothesis was that the part of the pitch movement that would by itself be enough to elicit the perception of a p-phrase would be the defining part specified by tonal targets, whereas any other normally occurring tonal movement would be an interpolation to or from neighbouring targets. Also the results of this experiment were in line with the assumption of H_pL_p phrase tones.

On the basis of these studies, the thesis thus suggests that...

1. ...there is no evidence for contrasting accents and instead the data presented here can be modelled in terms of a unified specification of phrase tones.

2. ...these phrase tones define a falling contour on the p-phrase level and should be transcribed as H_pL_p.

3. ...Finnish is a phrase language in the sense of Féry (2010).

Of central importance for these claims is the conclusion that peak timing did not show any systematic variation indicative of contrasting accents in the data used here. The following section will therefore review the evidence from the different chapters, before section 6.3 addresses the role of Finnish as a phrase language in prosodic typology.

6.2 Peak timing

In the data of this thesis, variation in peak timing occurred in relation to position, focus and quantity.

Positional differences were only investigated in Chapter 2, where they were correlated with grammatical function due to fixed SVO word order. For all three positions, a majority of peaks occurred on the first syllable vowel nucleus. The positions only differed in how strong the predominance of this alignment was, with the largest variation appearing on sentence-medial verbs. Chapter 2 accounted for this variation in terms of phrasing, concluding that verbs frequently formed p-phrases with the following objects. It found this approach to be superior over an accent-based alternative stating that finite verbs are unaccented unless narrowly focused (see Välimaa-Blum, 1993) in accounting for variable peak timing and duration patterns at the same time.

Focus conditions varied in the data of Chapter 2—re-evaluated in Chapter 3—and Chapter 4; additionally, a study of the effect of differences in peak timing on the perception of narrow focus appeared in section 3.3 of Chapter 3. Both Chapter 2 and Chapter 4 observed differences in peak timing for the focus conditions. Interestingly, peaks were generally later for narrow focus words in Production Experiment II (Chapter 4), but this finding was in direct contradiction to the results of Production Experiment I, where narrow focus words displayed overall earlier peaks (see section 2.4.4). However, Chapter 3 concluded that the differences between information structural roles and positions were not significant in a more carefully edited version of the data set from Production Experiment I. Moreover, neither earlier nor later peaks induced a higher number of narrow focus responses in the perception study reported in section 3.3. Although the importance of this small pilot study should not be overstated, it is conspicuous that variation in peak timing induced no significant effects, while peak height did.

In both the original data set used in Chapter 2 and the data from Production Experiment II, variation in the timing of pitch maxima was larger in broad focus (and unfocused given) constituents than in narrow focus condition. It is thus suggested here the variation is responsible for the apparent difference between the two overall very similar data sets, as well as for the spurious finding of information structural differences in peak timing within both data sets. This variation is seen here as largely coming from variation in phrasing. Chapters 2 and 3 have shown several examples of pitch maxima not corresponding to phonological targets, but being part of a fall from a preceding word's peak or a rise to a following word's peak (see especially section 3.2.1 for examples and an assessment of frequency). Chapter 2 also showed that while prototypical phrasing patterns could be identified for certain information structures, there was still some variation even in the short SVO sentences evaluated there. Additionally, even between different instantiations of a given phrasing, a certain amount of non-contrastive variation is expected for the realisation of phrase tones (see Féry, 2010, as an example, consider Figures 2.4 and 2.5 again).

Finally, quantity was varied as an experimental factor in Chapter 2 and Chapter 4. While Production Experiment I only varied the quantity of first syllable vowels systematically (CV.X vs. CVV.X), Production Experiment II induced a larger number of first syllable structures (CV.X, CVC.X, CVC_g.X, CVV.X, CVVC.X) as well as a systematic variation in second syllable vowel quantity and number of syllables. This thesis suggests that tonal cues of quantity in Finnish are a matter of duration determined by syllable weight, following the time line model (see O'Dell, 2003; Suomi, 2007a, 2009, and the summary in section 1.5). This model sees the phonological specification of tonal contours as uniform, but distributed across different segmental structures and durations. Thus, the relative alignment of tonal movements and segmental structures is understood as cuing segmental quantities. Under the prosodic analysis suggested here, the relevant tonal aspect is the realisation of the H_p tone. Relative to the beginning of the first syllable vowel nucleus, the distance of this tone was largely stable. In both production experiments, peaks appeared on average somewhat later in CV.X words than in words with longer first syllables. But like the apparent differences between focus conditions, the quantity contrast did not persist in the analysis of the more carefully edited data set in Chapter 3. Also for the data in Chapter 4, the tendency was not consistently significant. Furthermore, the differences in Production Experiment II were not binary. Rather, a continuum emerged where peak timing was later for words with shorter first syllables and earlier in words with longer first syllables, i.e. those that contained more or phonologically longer segments. Also earlier studies found tonal contours to be identical for CVV and CVC syllables (e.g. Vainio et al., 2006). This is in line with the present assumption that quantity cues arise from the interplay of a relatively stable tonal contour and duration as determined by segmental structure, instead of different tonal contours directly expressing long vs. short quantity of individual segments (i.e. tonal contour X being realised on segments with quantity 1 and tonal contour Y on segments with quantity 2). A more probable alternative would be tonal cuing of quantity through different tonal targets at the foot level, as suggested by Vainio et al. (2006), but note that while Chapter 4 showed some effects of first syllable weight, it provided practically no evidence for an effect of quantity changes in the second syllable (see section 4.4.4.1, especially Table 4.11).

On a whole, the evidence available so far only points to tonal contours cuing first syllables quantity differences and first syllables are also usually the location of H_p tones (analysed as accent peaks in previous literature).[1] In line with its empirical findings, this thesis thus concludes that the time line model is well suited to explain tonal quantity cues and that quantity differences do not lead to principled distinctions in peak timing.

6.3 Finnish as a phrase language

On the basis of the data presented in the preceding chapters, this thesis suggests that Finnish can productively be analysed as a phrase language in the sense of Féry (2010), i.e. as a language for which prosody is defined by phrasing and phrase tones.

The dissertation has argued that variation in key areas of grammar and differences that would be signalled by the choice of accent type in intonation languages like English are instead expressed through phrasing and other prosodic means in Finnish. Thus, Finnish would not seem to fit the classifi-

[1] Note also that most previous studies on tonal cues of quantity like Vainio et al. (2006), Suomi (2007a, 2009), Vainio et al. (2010) and Järvikivi et al. (2010) have exclusively varied quantity in first syllables.

cation as a intonation language. This group is crucially characterised by the fact that while locations for tonal events may be lexically fixed (prototypically on stressed syllables), the form of these tonal events is free and varies to express post-lexical distinctions. In contrast to the defining flexibility of intonation languages, Finnish has always been described as displaying rather little variation in terms of intonational contours, see for example Välimaa-Blum's (1988) inventory containing a single pitch accent and Suomi's (2005; 2007a; 2007b; 2009) studies on accent uniformity. Thus, although Finnish intonation has been described in terms of accents associated with stressed syllables, it differs fundamentally from languages like English or Dutch. What has been described as pitch cues of primary stress is analysed here as realisations of phrase tones (cf. e.g. Tuomainen et al., 1999, but also see Iivonen et al., 1998 and Suomi et al., 2003 who distinguish stress and accent).

One can thus argue that Finnish, like the Indian languages discussed by Féry (2010), is a phrase language, albeit one that has stress manifested by segmental lengthening (e.g. Suomi and Ylitalo, 2004). Since stress can be present or absent for other intonational typological types like tone languages (see Jun, 2005a), this conclusion might not be so surprising at all. Because Finnish stress is never contrastive, it might actually be fairly easy to explain even the durational effects as just an additional boundary phenomenon. In this context, it is interesting that the analysis in section 3.4 discovered no effect of stress on intensity. Although the implications of this post-hoc study should not be overstated, it fits in with descriptions of stress as being less central and less clearly phonetically marked in Finnish than in Germanic languages (e.g. Sadeniemi, 1949; Niemi, 1984; Iivonen, 1998; Ylitalo, 2009).

Crucially, the prosodic typological classification suggested by Féry (2010) describes the behaviour of a language at the sentence / phrase level. Thus Féry (2010) lists French as an intonation language and thereby in the same category as English, although a lexical level classification would take the absence of stress as decisive. In the same vein, what is crucial for the classification of Finnish as a phrase language is not the absolute absence of (systematic) variation in alignment or contrasting accents, but the dominance of phrase-level phenomena in shaping prosody. Therefore, although for instance Iivonen (1993, 1998) describes alternative accents differing from the default contour also in tonal alignment, these patterns are mostly relatively rare contours conveying stylistic or metalinguistic differences. The data of this thesis, displaying substantive prosodic variation due to manipulation of information structure, agree very well the the characterisation as a phrase language.

In analysing prosodic variation in terms of phrasing instead of accents, the dissertation has largely concentrated on the p-phrase level, but also discussed topics like syllable weight (Chapter 4) and provided some suggestions regarding the i-phrase, most prominently that it is aligned with the right edge of a focus constituent (Chapter 2). Table 6.1 gives an overview of

Table 6.1: Overview of the prosodic hierarchy employed here.

Prosodic unit	Correlates
Mora	duration, syllable weight
Syllable	inter alia relevant in durational adjustment, especially quantity
Prosodic word	vowel harmony, phonotactics, etc.
p-phrase	H_pL_p, (duration, also see Niemi, 1984)
i-phrase	duration, voice quality, H_i / L_i

the prosodic units that the analysis referred to. Notice that this analysis intended to account for the data of this thesis and makes no claim to completeness. In particular, issues like recursivity have not been addressed. More research into prosodic phrasing in Finnish is clearly necessary. For example, noun phrases were realised consistently as two p-phrases in Chapter 5, but often—though not always—as one p-phrase in Chapter 4. It is likely that variation in phrasing for noun phrases is possible and influenced by several factors that need to be explored further. This thesis suggests that the analysis as a phrase language provides a productive approach to intonation and phrasing in Finnish.

Chapter 7

References

Anttila, A. (2009). Derived environment effects in colloquial Helsinki Finnish. In Hanson, K. and Inkelas, S., editors, *The Nature of the Word: Essays in Honor of Paul Kiparsky*, pages 433–460. MIT Press, Cambridge, Massachusetts.

Anttila, A. (2010). Word stress in Finnish. Handout presented at Yale University, February 1, 2010.

Anttila, A. (2012). Modelling phonological variation. In Cohn, A. C., Fougeron, C., and Huffman, M. K., editors, *The Oxford handbook of laboratory phonology*, pages 76–91. Oxford University Press, Oxford.

Anttila, H. (2008). The effect of interrogative function on intonation in spontaneous and read Finnish. Master's thesis, University of Helsinki.

Arnhold, A., Vainio, M., Suni, A., and Järvikivi, J. (2010). Intonation of Finnish Verbs. In *Speech Prosody 2010*, pages 100054:1–4.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R.* Cambridge University Press.

Baayen, R. H., Davidson, D., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412.

Beckman, M. E. and Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3:255–309.

Boersma, P. and Weenink, D. (2010). Praat. Doing phonetics by Computer (Version 5.1.29).

Bruce, G. (1977). *Swedish word accents in sentence perspective.* Travaux de l' Institut de Linguistique de Lund 12. Gleerup, Lund.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Li, C. N., editor, *Subject and topic*, pages 27–55. Academic Press, New York.

Chen, A. (2011). What's in a rise: Evidence for an off-ramp analysis of Dutch intonation. *Proceedings of the ICPhS XVII, 17-21 August 2011, Hong Kong*, pages 448–451.

Chomsky, N. (1971). Deep structure, surface structure, and semantic interpretation. In Steinberg, D. D. and Jakobovits, L. A., editors, *Semantics. An interdisciplinary reader in philosophy, linguistics and psychology*, pages 183–216. Cambridge University Press, Cambridge.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English.* Studies in Language. Harper & Row, New York.

Cruttenden, A. (1997). *Intonation.* Cambridge University Press, Cambridge, 2nd edition.

Elenbaas, N. and Kager, R. (1999). Ternary rhythm and the lapse constraint. *Phonology*, 16(3):273–329.

Féry, C. (1993). *German intonational patterns.* Linguistische Arbeiten 285. Max Niemeyer Verlag, Tübingen.

Féry, C. (2010). Indian Languages as Intonational 'Phrase Languages'. In Hasnain, I. and Chaudhury, S., editors, *Problematizing Language Studies. Festschrift for Rama Agnihotri*, pages 288–312. Aakar Books, Delhi.

Féry, C. (2011). German sentence accents and embedded prosodic phrases. *Lingua*, 121(13):1906–1922.

Féry, C. and Ishihara, S. (2009). How Focus and Givenness Shapes Prosody. In Zimmermann, M. and Féry, C., editors, *Information structure from different perspectives*, pages 36–63. Oxford University Press, Oxford.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and speech*, 1(2):126–152.

Goldsmith, J. (1976). *Autosegmental Phonology.* PhD thesis, MIT.

Gordon, M. and Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, 29(4):383–406.

Grabe, E. (1998). *Comparative Intonational Phonology: English and German.* MPI Series 7. Ponsen en Looien, Wageningen.

Grice, M., Baumann, S., and Benzmüller, S. (2005). German intonation in autosegmental-metrical phonology. In Jun, S.-A., editor, *Prosodic typology: The phonology of intonation and phrasing*, pages 55–83. Oxford University Press, Oxford.

Gundel, J. K., Hedberg, N., and Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):274–307.

Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics*, 19(2):377–417.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Research surveys in linguistics. Cambridge University Press, Cambridge.

Gussenhoven, C. (2005). Transcription of Dutch intonation. In Jun, S.-A., editor, *Prosodic typology: The phonology of intonation and phrasing*, pages 118–145. Oxford University Press, Oxford.

Gussenhoven, C. (2008). Semantic judgments as evidence for the intonational structure of Dutch. *Proceedings of Speech Prosody 2008*, pages 297–300.

Hanson, K. and Kiparsky, P. (1996). A parametric theory of poetic meter. *Language*, 72(2):287–335.

Harrikari, H. (2000). *Segmental length in Finnish – Studies within a constraint-based approach*. Publications / University of Helsinki, Department of General Linguistics 33. University of Helsinki, Helsinki.

Harrikari, H. (2003). Quantity-sensitivity of syllabic trochees revisited. The case of dialectal gemination in Finnish. In van de Weijer, J., van Heuven, V. J., and van der Hulst, H., editors, *The phonological spectrum. Volume II: Suprasegmental Structure*, pages 111–126. Benjamins, Amsterdam.

Hayes, B. (1980). *A metrical theory of stress rules*. PhD thesis, MIT.

Hirsimäki, T., Creutz, M., Siivola, V., Kurimo, M., Virpioja, S., and Pylkkönen, J. (2006). Unlimited vocabulary speech recognition with morph language models applied to finnish. *Computer Speech & Language*, 20(4):515–541.

Hirvonen, P. (1970). *Finnish and English communicative intonation*. Publications of the Phonetics Department of the University of Turku 8. University of Turku, Turku.

Hyman, L. M. (2006). Word-prosodic typology. *Phonology*, 23(2):225–257.

Iivonen, A. (1978). Is there interrogative intonation in Finnish? In Gårding, E., Bruce, G., and Bannerts, B., editors, *Nordic prosody. Papers from a symposium*, pages 43–53, Lund. Lund University.

Iivonen, A. (1987). Paradigm of recurrent accent types in Finnish. In Gregersen, K. and Basbøll, H., editors, *Nordic Prosody IV. Papers from a symposium*, Odense University Studies in Linguistics Vol. 7, pages 115–126, Odense. Odense University Press.

Iivonen, A. (1993). Acoustical features of accentuation as a basis of prosodic style. In Granström, B. and Nord, L., editors, *Nordic Prosody VI. Papers from a symposium*, pages 99–126, Stockholm. Almqvist & Wiksell International.

Iivonen, A. (1998). Intonation in Finnish. In Hirst, D. and Di Cristo, A., editors, *Intonation Systems: A Survey of Twenty Languages*, pages 311–327. Cambridge University Press, Cambridge.

Iivonen, A. and Harnud, H. (2005). Acoustical comparison of the monophthong systems in Finnish, Mongolian and Udmurt. *Journal of the International Phonetic Association*, 35(1):59–71.

Iivonen, A., Nevalainen, T., Aulanko, R., and Kaskinen, H. (1987). *Puheen intonaatio*. Gaudeamus, Helsinki.

Iivonen, A., Niemi, T., and Paananen, M. (1998). Do F0 peaks coincide with lexical stress? A comparison of English, Finnish and German. In Werner, S., editor, *Nordic prosody: proceedings of the VIIth conference, Joensuu 1996*, pages 141–158, Frankfurt am Main. Peter Lang.

Iivonen, A. and Yli-Luukko, E. (1991). Sentence accentuation in Finnish dialects. A pilot study based on two dialect samples. In Aulanko, R. and Leiwo, M., editors, *Studies in logopedics and phonetics 2*, Publications of the Department of Phonetics, Series B: Phonetics, Logopedics and Speech Communications 3, pages 33–56. University of Helsinki, Helsinki.

Ito, J. and Mester, A. (2009). The extended prosodic word. In Kabak, B. and Grijzenhout, J., editors, *Phonological domains: Universals and derivations*, pages 135–194. Mouton de Gruyter, The Hague.

Järvikivi, J., Aalto, D., Aulanko, R., and Vainio, M. (2007). Perception of vowel length: Tonality cues categorization even in a quantity language. In *Proceedings of the 16th International Congress of Phonetic Sciences*, pages 693–696.

Järvikivi, J., Vainio, M., and Aalto, D. (2010). Real-Time Correlates of Phonological Quantity Reveal Unity of Tonal and Non-Tonal languages. *PLoS ONE*, 5(9):e12603.

Jokinen, K. (2005). A Grammar for Finnish Discourse Patterns. In Arppe, A., Carlson, L., Lindén, K., Piitulainen, J., Suominen, M., Vainio, M., Westerlund, H., and Yli-Jyrä, A., editors, *Inquiries into Words, Constraints and Contexts. Festschrift for Kimmo Koskenniemi on his 60th Birthday.*, CSLI Studies in Computational Linguistics ONLINE, pages 227–240. CSLI Publications, Stanford, CA.

Jun, S.-A. (2005a). Prosodic Typology. In Jun, S., editor, *Prosodic Typology. The Phonology of Intonation and Phrasing.*, pages 430–458. Oxford University Press, Oxford.

Jun, S.-A. (2005b). *Prosodic typology: The phonology of intonation and phrasing.* Oxford University Press, Oxford.

Karlsson, F. (1977). Morphotactic structure and word cohesion in Finnish. In Sajavaara, K. and Lehtonen, J., editors, *Contrastive papers*, Reports from the Department of English 4, Jyväskylä contrastive studies 4, pages 59–74. University of Jyväskylä, Jyväskylä.

Karlsson, F. (1983). *Suomen kielen äänne- ja muotorakenne.* Werner Sönderström Osakeyhtiö, Porvoo/Helsinki/Juva.

Karlsson, F. (1999). *Finnish: An Essential Grammar.* Routledge, London/New York.

Karlsson, F. (2005). Phonotactic Complexity of Finnish Nouns. In Arppe, A., Carlson, L., Lindén, K., Piitulainen, J., Suominen, M., Vainio, M., Westerlund, H., and Yli-Jyrä, A., editors, *Inquiries into Words, Constraints, and Contexts. Festschrift for Kimmo Koskenniemi on his 60th Birthday*, CSLI Studies in Computational Linguistics ONLINE, pages 65–70. CSLI Publications, Stanford, California.

Karlsson, F. (2006). Kommentti Kari Suomelle. *Virittäjä*, 110:264.

Karlsson, F. (2007). Morayleistyksen rajoitukset. *Virittäjä*, 110:99–100.

Karlsson, F. and Lehtonen, J. (1977). *Alkukahdennus. Näkökohtia eräistä suomen kielen sandhi-ilmiöistä.* Publications of the Department of Finnish and General Linguistics of the University of Turku. Turku University, Turku.

Karttunen, L. (2006). The insufficiency of paper-and-pencil linguistics: the case of Finnish prosody. *Rutgers Optimality Archive*, ROA-818:3–16.

Karvonen, D. H. (2005). *Word Prosody in Finnish.* PhD thesis, University of California, Santa Cruz.

Krifka, M. (1984). Fokus, Topik, syntaktische Struktur und semantische Interpretation. Unpublished manuscript, available online: `http://amor.cms.hu-berlin.de/~{}h2816i3x/Publications/Krifka%201984%20Fokus.PDF`.

Krifka, M. (2007). Basic notions of information structure. In Féry, C., Fanselow, G., and Krifka, M., editors, *The notions of information structure*, Working Papers of the SFB632, Interdisciplinary Studies on Information Structure (ISIS) 6, pages 13–56. Universitätsverlag Potsdam, Potsdam.

Ladd, D. R. (1996). *Intonational Phonology.* Cambridge Studies in Linguistics 79. Cambridge University Press, Cambridge.

Ladd, D. R., Mennen, I., and Schepman, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America*, 107:2685–2696.

Lambrecht, K. (1994). *Information structure and sentence form. Topic, focus and the mental representations of discourse referents.* Cambridge studies in linguistics 71. Cambridge University Press, Cambridge.

Leben, W. R. (1971). The Morphophonemics of Tone in Hausa. In Kim, C.-W. and Stahlke, H., editors, *Papers in African Linguistics*, pages 201–218. Linguistic Research, Inc., Edmonton, Alberta.

Lehiste, I. (1970). *Suprasegmentals.* MIT Press, Cambridge.

Lehtonen, J. (1970). *Aspects of Quantity in Standard Finnish.* Studia Philologica Jyväskyläensia VI. Gummerus, Jyväskylä.

Liberman, M. (1975). *The Intonational System of English.* PhD thesis, MIT.

Liberman, M. and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic inquiry*, 8(2):249–336.

Malmberg, B. (1949). Review of "Voyelles longues et voyelles brèves." by Marguerite Durand. *Studia linguistica*, 3(1):39–59.

Mixdorff, H., Vainio, M., Werner, S., and Järvikivi, J. (2002). The manifestation of linguistic information in prosodic features of Finnish. In *Proceedings of Speech Prosody 2002*, pages 511–514, Aix-en-Provence.

Nakai, S., Kunnari, S., Turk, A., Suomi, K., and Ylitalo, R. (2009). Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics*, 37(1):29–45.

Nespor, M. and Vogel, I. (1986). *Prosodic Phonology.* Foris Publications, Dordrecht.

Niemi, J. (1984). *Word level stress and prominence in Finnish and English. Acoustic experiments on production and perception.* University of Joensuu, Joensuu.

O'Connor, J. D. and Arnold, G. F. (1973). *Intonation of Colloquial English.* Longman, London.

O'Dell, M. (2003). *Intrinsic timing and quantity in Finnish.* Acta Universitatis Tamperensis 979. Tampere University Press, Tampere.

O'Dell, M. L., Lennes, M., Werner, S., and Nieminen, T. (2007). Looking for rhythms in conversational speech. In Trouvain, J. and Barry, W. J., editors, *Proceedings of the 16th International Congress of Phonetic Sciences*, pages 1201–1204, Saarbrücken. Universität des Saarlandes.

Ogden, R. and Routarinne, S. (2005). The communicative function of final rises in Finnish intonation. *Phonetica*, 62(2-4):160–175.

Okobi, A. O. (2006). *Acoustic Correlates of Word Stress in American English.* PhD thesis, MIT.

Paunonen, H. (1995). *Suomen kieli Helsingissä. Huomioita Helsingin puhekielen historiallisesta taustasta ja nykyvariaatiosta.* University of Helsinki, Helsinki.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation.* PhD thesis, MIT.

Pierrehumbert, J. B. and Beckman, M. E. (1988). *Japanese Tone Structure.* Linguistic Inquiry Monograph 15. MIT Press, Cambridge.

Prieto, P., Shih, C., and Nibert, H. (1996). Pitch downtrend in Spanish. *Journal of Phonetics*, 24(4):445–473.

R Development Core Team (2010). R: A Language and Environment for Statistical Computing.

Raitio, T., Suni, A., Yamagishi, J., Pulakka, H., Nurminen, J., Vainio, M., and Alku, P. (2011). Hmm-based speech synthesis utilizing glottal inverse filtering. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(1):153–165.

Ratcliff, R. (1993). Methods for Dealing With Reaction Time Outliers. *Psychological Bulletin*, 114(3):510–532.

Reinhart, T. (1981). Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, 27(1):53–94.

Rooth, M. (1985). *Association with Focus.* PhD thesis, University of Massachusetts, Amherst.

Rooth, M. (1992). A theory of focus interpretation. *Natural language semantics*, 1(1):75–116.

Sadeniemi, M. (1949). *Metriikkamme perusteet ja sovellutusta moderneihin ja antiikin mittoihin*. Suomalaisen Kirjallisuuden Seuran Toimituksia, Helsinki.

Sadeniemi, M. (1965). Sanapaino ja lausepaino. *Virittäjä*, 69:101–102.

Schepman, A., Lickley, R., and Ladd, D. R. (2006). Effects of vowel length and "right context" on the alignment of Dutch nuclear accents. *Journal of Phonetics*, 34(1):1–28.

Selkirk, E. O. (1980). The role of prosodic categories in English word stress. *Linguistic inquiry*, 11(3):563–606.

Selkirk, E. O. (1981). On prosodic structure and its relation to syntactic structure. In Fretheim, T., editor, *Nordic Prosody II. Papers from a symposium*, pages 111–140, Trondheim. Tapir.

Selkirk, E. O. (2000). The interaction of constraints on prosodic phrasing. In Horne, M., editor, *Prosody: Theory and Experiment. Studies presented to Gösta Bruce*, pages 231–261. Kluwer, Dordrecht, Boston and London.

Selkirk, E. O. (2007). Contrastive Focus, Givenness and the Unmarked Status of "Discourse-New". In Féry, C., Fanselow, G., and Krifka, M., editors, *The notions of information structure*, Working Papers of the SFB632, Interdisciplinary Studies on Information Structure (ISIS) 6, pages 125–145. Universitätsverlag Potsdam, Potsdam.

Sulkala, H. and Karjalainen, M. (1992). *Finnish*. Routledge, London / New York.

Suomi, K. (1985). On detecting words and word boundaries in Finnish: A survey of potential word boundary signals. *Nordic Journal of Linguistics*, 8(02):211–231.

Suomi, K. (2004). Moraic patterns in Finnish prosody and lexicon. In Peltola, M. S. and Tuomainen, J., editors, *Studies in speech communication. Dedicated to Ilkka Raimo on his 65th birthday*, Publications of the Department of Finnish and General Linguistics of the University of Turku 72, pages 87–97. University of Turku, Turku.

Suomi, K. (2005). Temporal conspiracies for a tonal end: segmental durations and accentual f0 movement in a quantity language. *Journal of Phonetics*, 33:291–309.

Suomi, K. (2007a). Accentual tonal targets and speaking rate in Northern Finnish. In Granström, B., Hjalmarsson, A., House, D., and Wik, P., editors, *Proceedings of Fonetik, TMH-QPSR 50*, pages 109–112, Stockholm. Department of Speech, Music and Hearing, KTH.

Suomi, K. (2007b). On the tonal and temporal domains of accent in Finnish. *Journal of Phonetics*, 35:40–55.

Suomi, K. (2009). Durational elasticity for accentual purposes in Northern Finnish. *Journal of Phonetics*, 37(4):397–416.

Suomi, K., Toivanen, J., and Ylitalo, R. (2003). Durational and Tonal Correlates of Accent in Finnish. *Journal of Phonetics*, 31:113–138.

Suomi, K., Toivanen, J., and Ylitalo, R. (2008). *Finnish Sound Structure. Phonetics, Phonology, Phonotactics and Prosody.* Studia Humaniora Ouluensia 9. University of Oulu, Oulu.

Suomi, K. and Ylitalo, R. (2004). On Durational Correlates of Word Stress in Finnish. *Journal of Phonetics*, 32(1):35–63.

Tuomainen, J., Werner, S., Vroomen, J., and de Gelder, B. (1999). Fundamental frequency is an important acoustic cue to word boundaries in spoken Finnish. In Ohala, J. J., Hasegawa, Y., Ohala, M., Granville, D., and Bailey, A. C., editors, *Proceedings of the 14th International Congress of Phonetic Sciences, August 1-7, 1999, San Francisco, USA*, pages 921–923.

Vainio, M., Aalto, D., Järvikivi, J., and Suni, A. (2006). Quantity and tone in Finnish lexically stressed syllables. In Belotel-Grenié, A. and Grenié, M., editors, *Proceedings of the 2nd International Symposium on Tonal Aspects of Languages*, pages 121–124, La Rochelle. Université de La Rochelle.

Vainio, M. and Järvikivi, J. (2006). Tonal Features, Intensity, and Word Order in the Perception of Prominence. *Journal of Phonetics*, 34:319–342.

Vainio, M. and Järvikivi, J. (2007). Focus in production: Tonal shape, intensity and word order. *The Journal of the Acoustical Society of America*, 121(2):EL55–EL61.

Vainio, M., Järvikivi, J., Aalto, D., and Suni, A. (2010). Phonetic tone signals phonological quantity and word structure. *Journal of the Acoustic Society of America*, 128(3):1313–1321.

Vainio, M., Mixdorff, H., Järvikivi, J., and Werner, S. (2003). The Production and Perception of Focus in Finnish. In *Proceedings of the 15th International Congress of Phonetic Sciences*, pages 1831–1834.

Välimaa-Blum, R. (1988). *Finnish existential clauses - their syntax, pragmatics and intonation.* PhD thesis, The Ohio State University.

Välimaa-Blum, R. (1993). A Pitch Accent Analysis of Intonation in Finnish. *Ural-Altaische Jahrbücher*, 12:82–94.

Valldubí, E. and Vilkuna, M. (1998). On rheme and kontrast. *Syntax and Semantics*, 29:79–108.

Vilkuna, M. (1989). *Free Word Order in Finnish: Its Syntax and Discourse Functions.* Suomalaisen Kirjallisuuden Seuran Toimituksia, Helsinki.

Vilkuna, M. (1995). Discourse Configurationality in Finnish. In É. Kiss, K., editor, *Discourse Configurational Languages*, pages 244–268. Oxford University Press, New York/Oxford.

Wiik, K. (1975). On vowel duration in Finnish dialects. In Hallap, V., editor, *Congressus Tertius Internationalis Fenno-Ugristarum Tallinnae habitus 17.–23. VIII 1970. Pars I. Acta Linguistica*, pages 415–424, Tallinn. Valgus.

Wiik, K. (1977). Suomen tavuista. *Virittäjä*, 81:265–278.

Wiik, K. (1981). *Fonetiikan perusteet.* Werner Sönderström Osakeyhtiö, Porvoo/Helsinki/Juva.

Wiik, K. (1985). Suomen murteiden vokaalien kestoista. In Aaltonen, O. and Hulkko, T., editors, *Fonetiikan päivät — Turku 1985. XIII Fonetiikan päivillä Turun yliopistossa 30.–31.8.1985 pidetyt esitelmät*, pages 253–317. University of Turku.

Wiik, K. (1988). F0 huipun sijainti suomessa. In Karjalainen, M. and Laine, U. K., editors, *Fonetiikan päivät — Espoo 1988. Papers from the 15th meeting of Finnish phoneticians*, Publications 31, Otaniemi. Helsinki University of Technology, Faculty of Electrical Engineering, Acoustics Laboratory.

Wiik, K. and Lehiste, I. (1968). Vowel quantity in Finnish disyllabic words. In Ravila, P., editor, *Congressus Secundus Internationalis Fenno-Ugristarum, Helsingiae habitus 23–28. VIII. 1965. Pars I*, pages 569–574, Helsinki. Societas Fenno-Ugrica/Suomalais-Ugrilainen Seura.

Williams, E. (1976). Underlying Tone in Margi and Igbo. *Linguistic Inquiry*, 7(3):463–484.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0contours. *Journal of Phonetics*, 27:55–105.

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 46(3-4):220–251.

Ylitalo, R. (2009). *The realisation of prominence in three varieties of standard spoken Finnish.* Acta Universitatis Oluensis, B Humanoiara 88. University of Oulu, Oulu.

Appendix A

Background information on participants of Production Experiment I

This appendix provides information about the subjects participating in the production experiment discussed in Chapter 2.

Table A.1: Age, sex, native language and foreign languages spoken by participants of Production Experiment I.

Id	Age	Sex	Native language	Other languages spoken
s01	24	f	Finnish	English, Swedish, French, Spanish, Estonian
s02	28	f	Finnish	English, Swedish, French, Spanish, Swahili
s03	25	f	Finnish	Swedish, English, French, Spanish, Nepali
s04	31	m	Finnish, Russian	English, Swedish, French, Greek
s05	25	f	Finnish	English, French, Swedish, Nepali, Swahili, Spanish, Italian, Russian
s06	35	m	Finnish	Swedish, English, German, (French, Latin)
s07	20	f	Finnish	English, Swedish, German, Spanish, French
s08	27	f	Finnish	Spanish, Portuguese, English, Swedish
s09	21	f	Finnish	English, German, Swedish, Spanish
s10	23	m	Finnish	Swedish, English, German, Spanish, Russian
s11	24	m	Finnish	English
s12	25	f	Finnish	German, English, Swedish, Spanish, Dutch
s13	25	f	Finnish	English, Russian
s14	22	f	Finnish, (Bulgarian)	English, French, Swedish, (Norwegian, Danish)
s15	26	m	Finnish	Swedish, English, French, German, Latin, Greek
s16	28	m	Finnish	Swedish, English, French, Russian, Lithuanian
s17	25	m	Finnish	Swedish, English, German, French, Spanish
s18	22	f	Finnish	Swedish, English, French, Japanese

Table A.2: Dialect and places of residency of participants of Production Experiment I.

Id	Dialect	Lived in
s01	standard	Helsinki
s02	standard	Nummijärvi, Tampere, Spain, Helsinki
s03	standard, Helsinki dialect (stadin slangi)	Helsinki, Riga, Rovaniemi, Stockholm
s04	mostly standard, sometimes North Karelian dialect	Helsinki, Brussels, Moscow, Switzerland, Joensuu, Greece
s05	standard	Helsinki, France, Pori, Nepal
s06	standard	Imatra, Äänekoski, Espoo, Helsinki, London, Norwich
s07	standard	Helsinki, Australia, Espoo
s08	standard	Espoo, Karjaa, Turku, Helsinki, Vantaa, Spain, Portugal
s09	standard	Vantaa, Klaukkala, Helsinki
s10	standard, Helsinki dialect	Helsinki, Espoo
s11	standard	Tammisaari, Helsinki
s12	standard	Espoo, Mäntsälä, Helsinki, Scotland, Germany
s13	standard	Helsinki, Kouvola
s14	standard	Helsinki
s15	standard	Kerava, Helsinki
s16	standard	Espoo, Helsinki, Vekaranjärvi, Hamina, Sheffield, Vilnius
s17	standard	Helsinki, Kirkkonummi, Espoo
s18	standard	Espoo

Table A.3: Information about possible speech pathologies and factors affecting speech physiology for participants of Production Experiment I.

Id	Speech pathology	Cold	Smoking
s01	s-problem as 7-year-old, corrected	no	no
s02	no	no	stopped
s03	no	no, but allergy	occasionally
s04	no	yes, larynx inflammation	stopped
s05	no	recently	no
s06	no	no	no
s07	no	recently	occasionally
s08	no	no	no
s09	no	no	no
s10	no, r-problem as a child	recently	occasionally
s11	no	no	no
s12	no	no	no
s13	no	no	no
s14	no	no, but allergy recently	no
s15	no	no	no
s16	no	no	no
s17	no	no	no
s18	no	no	no

Appendix B

Vowel sequences and diphthongs in the materials of Production Experiment I

The materials of Production Experiment I were generally designed to include only disyllabic words (cf. 2.2.1, list of sentences repeated in Table B.1). However, they contain three words that would generally be considered trisyllabic, since the vowel sequences *ea* [eɑ], *yä* [yæ] and *ua* [uɑ] are not diphthongs (Wiik, 1977; Karlsson, 1983). The three words, *loi.me.a* 'blanket', *lyi.jy.ä* 'lead' and *le.lu.a* 'toy', all were sentence-final objects. These items were nevertheless analysed together with the rest of the elicited data and generally treated as if they were disyllabic, i.e. syllabified *loi.mea*, *lyi.jyä* and *le.lu.a* (also see Suomi et al., 2008, p. 50, on ambivalent judgements regarding diphthongs and syllable boundaries).

Table B.1: Lexical material used in Production Experiment I.

	Subject	Verb	Object	Translation
1	Niilo	maalaa	liinaa	"Niilo (m) paints (a) cloth."
2	Moona	liimaa	naavaa	"Moona (f) glues lichen."
3	Maini	neuloi	loimea	"Maini (f) knitted a blanket."
4	Väinö	jauhoi	lyijyä	"Väinö (m) ground lead."
5	Jani	töni	lavaa	"Jani (m) pushed a platform."
6	Jimi	luki	menua	"Jimi (m) read the menu."
7	Manu	hali	lelua	"Manu (m) hugged a toy."
8	Jali	nyki	lanaa	"Jali (m) tugged a levelling drag."

In editing the materials, the distinction between disyllabic and trisyllabic object items was difficult to make from a phonetic point of view. The syllable boundary is for example not marked with a glottal stop, which according to Karlsson (1977) precedes vowel-initial syllables only when they are word-initial. Even if such a regularity existed, it would have been difficult to detect for the current cases of sentence-final syllables, which were mostly realised with creaky voice (cf. section 2.4.5). Also with respect to the formant structure, no clear indicators were found. Overall, the boundaries between the two vowels in the combinations [eɑ], [yæ] and [uɑ] generally seemed equally difficult to locate as those between the two parts of the diphthongs [ɑi], [eu], [oi], [æi] and [yi].

Additionally, speakers sometimes applied the optional rule of vowel coalescence. Anttila (2009) describes this phenomenon of colloquial Helsinki Finnish in the following way: Two heterosyllabic, not primary stress carrying vowels, of which the second one is [+ low], are realised as one long vowel (also see Paunonen, 1995, pp. 106–150). For example, *loimea* was sometimes pronounced as [loi.meː] instead of [loi.me.ɑ] in the current materials. In these instances, the resulting realisation is clearly disyllabic.

In terms of the duration of what was labelled as the second syllable here, no clear differences between truly disyllabic and phonologically trisyllabic words emerged. For the quantity 2 object items, a Kolmogorov-Smirnov test indicated that the durations of the strings [me.ɑ] and [jy.æ] (in *loimea* and *lyijä*) were part of the same distribution as the syllables [nɑː] and [ʋɑː] (in *liinaa* and *naavaa*) ($D = 0.05$, $p > .92$). A paired t-test over speaker means supported the conclusion that the difference was not significant ($t(16) = -0.29$, $p > .78$). For the quantity 1 objects, the Kolmogorov-Smirnov test indicated two different distributions ($D = 0.20$, $p < .001$). However, the mean duration of the 'second syllable' was clearly longer for the words *lavaa* and *lanaa* (mean 281 ms) than for *menua* and *lelua* (mean 269 ms), contrary to what would be expected if the latter two contain an additional syllable boundary.

Appendix C

Treatment of very small pitch movements in the materials of Production Experiment I

This section details the policy for editing the data from Production Experiment I with respect to very small pitch movements. The aim of this editing was to sort out imperceptible tonal movements from some, but not all analyses.

In the automatic classification of tonal contours as 'LHL', 'HL', 'LH' or 'flat', the difference between two neighbouring f_0 turning points within a word (i.e. between L1 and H and between H and L2) was evaluated as significant when its magnitude was at least 0.8 st. Likewise, f_0 measurement points were disregarded in the analysis of alignment (see section 2.4.4) when the measured pitch differed by less than 0.8 st from that of the neighbouring point(s) within the same word. In contrast, values differing less than 0.8 st from their neighbours were not discarded from the analysis of pitch scaling (see section 2.4.2), since excluding these turning points would have eliminated measurements from words with a small pitch range. This would have distorted the data and possibly hampered the discovery of pitch range compression strategies.

The basis for the 0.8 st criterion was an observation by Fry (1958). He found that a difference in the pitch level of two adjacent syllables leads to the perception of stress. This perception showed an all-or-none effect, i.e. it generally arose irrespective of the size of the step. However, the minimal pitch difference that Fry used in his materials was a step of 5 Hz (from a baseline of 97 Hz, i.e. 0.8 st). A preliminary experiment with 3 Hz-steps had led to a dispersion of the listeners' judgements, leading Fry to conclude that a pitch difference of this size "may well have been too small to cause the all-or-none effect in perception of stress" (Fry, 1958, p. 144). Alternative cut-off criteria are of course possible. For example, Iivonen et al. (1998) use a just noticeable difference criterion of 1 st for determining whether a syllable carries an f_0 peak. Here, 0.8 st was adopted to provide minimal trimming of the data.

Appendix D

Segmental alignment in the materials of Production Experiment I

This appendix provides data on segmental alignment of the three f_0 points—L1, H and L2—measured on each word in the materials of of Production Experiment I. Since the materials consisted of CVCV words—where V represents a short vowel, long vowel or a diphthong—the f_0 measurement points were detected on one of four segments: The onset consonant of the first syllable (C1), the vowel nucleus of the first syllable (V1), the onset consonant of the second syllable (C2) or the vowel nucleus of the second syllable (V2). The following tables give the alignment of L1, H and L2 points to these consonant and vowel segments in absolute numbers (Table D.1) and in percent (Table D.2).

Table D.1: Alignment of f_0 turning points to segments in absolute number of occurrences for all experimental conditions.

		Position 1				Position 2				Position 3			
L1		C1	V1	C2	V2	C1	V1	C2	V2	C1	V1	C2	V2
Broad focus	Quantity 1	36	7	0	0	26	15	1	0	38	1	1	1
	Quantity 2	26	19	0	0	35	9	2	0	38	3	0	1
Narrow focus (CF)	Quantity 1	50	10	0	0	48	3	0	0	66	0	0	0
	Quantity 2	45	6	0	0	68	0	0	0	65	0	0	0
Narrow focus (IF)	Quantity 1	41	7	1	1	46	3	0	0	59	0	0	0
	Quantity 2	44	6	0	0	67	0	0	0	66	0	0	0
Unfocused (CF)	Quantity 1	58	14	4	0	22	22	4	1	15	13	5	1
	Quantity 2	49	21	2	0	29	26	6	2	29	13	5	1
Unfocused (IF)	Quantity 1	57	12	1	0	26	23	8	0	22	8	5	2
	Quantity 2	54	17	1	0	39	17	8	0	41	21	5	1
H		C1	V1	C2	V2	C1	V1	C2	V2	C1	V1	C2	V2
Broad focus	Quantity 1	14	34	13	6	12	16	5	26	17	43	0	3
	Quantity 2	3	48	0	15	17	21	3	20	23	37	0	3
Narrow focus (CF)	Quantity 1	3	52	10	3	2	63	3	0	0	67	1	0
	Quantity 2	5	60	0	1	4	64	0	0	1	66	0	0
Narrow focus (IF)	Quantity 1	7	46	7	6	5	59	2	1	9	59	0	0
	Quantity 2	4	63	0	1	5	62	0	0	4	64	0	0
Unfocused (CF)	Quantity 1	28	40	38	20	37	21	3	35	18	9	2	25
	Quantity 2	23	97	3	6	49	15	7	33	15	20	2	30
Unfocused (IF)	Quantity 1	22	60	28	15	28	16	5	43	31	12	1	21
	Quantity 2	20	100	3	5	49	10	10	42	18	30	5	29
L2		C1	V1	C2	V2	C1	V1	C2	V2	C1	V1	C2	V2
Broad focus	Quantity 1	0	3	2	56	0	5	7	28	0	12	25	23
	Quantity 2	0	4	6	41	1	6	6	24	2	43	11	4
Narrow focus (CF)	Quantity 1	0	1	5	61	0	3	5	52	0	3	31	33
	Quantity 2	0	5	12	47	0	22	13	28	0	51	13	3
Narrow focus (IF)	Quantity 1	0	0	10	50	0	6	3	49	0	9	31	27
	Quantity 2	0	8	12	46	0	20	14	31	0	58	6	4
Unfocused (CF)	Quantity 1	0	0	4	108	2	18	12	45	0	11	5	14
	Quantity 2	0	5	20	95	2	17	12	38	0	26	3	12
Unfocused (IF)	Quantity 1	0	0	7	108	0	10	13	41	2	11	14	16
	Quantity 2	0	2	11	107	0	14	8	47	1	30	4	10

Table D.2: Alignment of f_0 turning points to segments in percent, rounded, for all experimental conditions.

		Position 1				Position 2				Position 3			
L1		C1	V1	C2	V2	C1	V1	C2	V2	C1	V1	C2	V2
Broad focus	Quantity 1	84	16	0	0	62	36	2	0	93	2	2	2
	Quantity 2	58	42	0	0	76	20	4	0	90	7	0	2
Narrow focus (CF)	Quantity 1	83	17	0	0	94	6	0	0	100	0	0	0
	Quantity 2	88	12	0	0	100	0	0	0	100	0	0	0
Narrow focus (IF)	Quantity 1	82	14	2	2	94	6	0	0	100	0	0	0
	Quantity 2	88	12	0	0	100	0	0	0	100	0	0	0
Unfocused (CF)	Quantity 1	76	18	5	0	45	45	8	2	44	38	15	3
	Quantity 2	68	29	3	0	46	41	10	3	60	27	10	2
Unfocused (IF)	Quantity 1	81	17	1	0	46	40	14	0	59	22	14	5
	Quantity 2	75	24	1	0	61	27	12	0	60	31	7	1
H		C1	V1	C2	V2	C1	V1	C2	V2	C1	V1	C2	V2
Broad focus	Quantity 1	21	51	19	9	20	27	8	44	27	68	0	5
	Quantity 2	5	73	0	23	28	34	5	33	37	59	0	5
Narrow focus (CF)	Quantity 1	4	76	15	4	3	93	4	0	0	99	1	0
	Quantity 2	8	91	0	2	6	94	0	0	1	99	0	0
Narrow focus (IF)	Quantity 1	11	70	11	9	7	88	3	1	13	87	0	0
	Quantity 2	6	93	0	1	7	93	0	0	6	94	0	0
Unfocused (CF)	Quantity 1	22	32	30	16	39	22	3	36	33	17	4	46
	Quantity 2	18	75	2	5	47	14	7	32	22	30	3	45
Unfocused (IF)	Quantity 1	18	48	22	12	30	17	5	47	48	18	2	32
	Quantity 2	16	78	2	4	44	9	9	38	22	37	6	35
L2		C1	V1	C2	V2	C1	V1	C2	V2	C1	V1	C2	V2
Broad focus	Quantity 1	0	5	3	92	0	12	18	70	0	20	42	38
	Quantity 2	0	8	12	80	3	16	16	65	3	72	18	7
Narrow focus (CF)	Quantity 1	0	1	7	91	0	5	8	87	0	4	46	49
	Quantity 2	0	8	19	73	0	35	21	44	0	76	19	4
Narrow focus (IF)	Quantity 1	0	0	17	83	0	10	5	84	0	13	46	40
	Quantity 2	0	12	18	70	0	31	22	48	0	85	9	6
Unfocused (CF)	Quantity 1	0	0	4	96	3	23	16	58	0	37	17	47
	Quantity 2	0	4	17	79	3	25	17	55	0	63	7	29
Unfocused (IF)	Quantity 1	0	0	6	94	0	16	20	64	5	26	33	37
	Quantity 2	0	2	9	89	0	20	12	68	2	67	9	22

Appendix E

R code used for editing the data set from production experiment I before the statistical analysis of peak alignment

This section presents a commented version of the R code which sorted out measurements of f_0 maxima (H) that were not 'real peaks'. It thereby proceeded in two steps: First, it identified the relevant local f_0 minima, which were localised either on the same word as the H point or on the preceding or following word. Second, it determined whether the each measured H could remain in the data set or had to be sorted out because...

- ...it was not an f_0 turning point, but part of a pitch fall from the preceding word's peak ('fall from', possible for sentence-medial verbs and sentence-final objects).
- ...it was not an f_0 turning point, but part of a pitch rise to the following word's peak ('rise to', possible for sentence-initial subjects and sentence-medial verbs).
- ...it differed from the relevant preceding and following f_0 minima by less than 0.8 st ('flat', possible in all three positions).

The code for these two steps is rendered in sections E.1 and E.2, respectively. It is exemplified only for the sentence-medial verbs, which constituted the most complex case since they could exhibit both 'fall from' and 'rise to' contours. In this, the unedited data set is referred to as `allUnEd`. This set contained the originally measured three f_0 values per word—L1, H and

L2—in columns named vL1pitchST, vHpitchST and vL2pitchST for measurements from the verbs (values for subjects and objects were collected in separate columns, with each row containing values from one sentence). The timing of these points relative to the duration of the word was collected in columns entitled vL1relT, sHrelT and vL2relT. These timing values ranged from 0 to 1, with a value of 0.5 indicating that the respective point was realised exactly in the middle of the word. The values resulting from the evaluations were written into new columns called vL1Real and vL2Real for the minima and vRealH for the maxima. Note that the latter column contained f_0 values. The analysis of alignment considered the timing of H points that had a numerical value instead of an NA as their 'real' pitch.

E.1 Determining the relevant L values

For the verbs, the relevant minimum preceding a verb's f_0 maximum could be the originally detected L1 value or it could be located on the preceding subject. This could be the case when the subject had been realised completely with its pitch rising to the verb's peak or when the subject did carry an own contour, but the rise to the verb's peak started from the end of the fall from the subject's peak without an intervening plateau.

Similarly, the rightwards located f_0 minimum to which a verb's peak should be compared later could be on the verb (the original L2) or on the object—either when it carried a 'fall from' contour, in which case its L2 would be the relevant point, or when it did not, so that its L1 should be chosen for comparison.

The general approach of the code given below is to first identify all cases where the relevant minima would not be located on the verb according to objective criteria. When these criteria were not met, the originally measured L values were retained as the default.

E.1.1 Minima preceding the peaks

First, words with no or only an extremely short stretch of measurable f_0 (registered as vHpitchST) were identified. For those, the value for the 'real L1' was again put to 'not determinable' (NA).

```
allUnEd$vL1Real <- as.numeric(ifelse(is.na(allUnEd$vHpitchST), NA,
```

When the preceding word had no or only a minimal stretch of measurable pitch, the beginning of the rise could only be the L1 originally identified on the word itself (vL1pitchST).

```
   ifelse(is.na(allUnEd$sHpitchST), allUnEd$vL1pitchST,
   ifelse(is.na(allUnEd$sL1pitchST) & is.na(allUnEd$sL2pitchST),
     allUnEd$vL1pitchST,
```

Also for words preceded by a pause, it was assumed that the 'real L1' would not be on the preceding word.

```
ifelse(allUnEd$pause1 == "yes", allUnEd$vL1pitchST,
```

When neither L1 of the verb nor L2 of the subject was measurable originally, the code checked for the presence of a 'rise to' contour on the subject by determining whether the maximum of the subject was late in the word and had an f_0 value lower than the maximum of the verb. When both criteria were met, the beginning of the rise to the verb's peak was assumed to be the point originally identified as L1 of the subject.

```
ifelse(is.na(allUnEd$vL1pitchST) & is.na(allUnEd$sL2pitchST)
  & (allUnEd$sHrelT > (1/3)*2)
  & (allUnEd$sHpitchST < allUnEd$vHpitchST),
  allUnEd$sL1pitchST,
```

When only L2 of the subject was not measurable, the automatic process additionally made sure that the difference between the rise from the verb's L1 to its H was less than 0.8 st larger than the rise from the subject's H to the verb's H before classifying the subject as 'rise to' and its L1 as the beginning of the rise to the verb. The purpose was to ensure that no significant valley intervened between the pitch maximum of the subject and that of the verb. Apart from this, the following line of code is the same as the previous one.

```
ifelse(is.na(allUnEd$sL2pitchST) & (allUnEd$sHrelT > (1/3)*2)
  & (allUnEd$sHpitchST < allUnEd$vHpitchST)
  & ((allUnEd$vHpitchST - allUnEd$vL1pitchST)
  - (allUnEd$vHpitchST - allUnEd$sHpitchST)) < 0.8,
  allUnEd$sL1pitchST,
```

Since the possibility of a rise starting on the subject had thus been checked for all cases where the subject's L2 was not measurable, the originally measured L1 of the verb was assumed to correspond to the real start of the rise in all other cases of missing values for the subject's L2.

```
ifelse(is.na(allUnEd$sL2pitchST), allUnEd$vL1pitchST,
```

Next, the code checked the occurrence of a 'rise to' contour on the subject when L1 of the verb had not been measurable (but L2 of the subject had been measured originally). The procedure was the same as above, but the following line additionally ascertained that the fall from the subject's H to its L2 was smaller than 0.8 st and did thus not constitute a relevant pitch fall disturbing the rising movement. When all criteria were met, the 'real L1' of the verb was taken to be the point originally identified as L1 of the subject.

```
ifelse(is.na(allUnEd$vL1pitchST) & (allUnEd$sHrelT > (1/3)*2)
  & (allUnEd$sHpitchST < allUnEd$vHpitchST)
  & abs(allUnEd$sHpitchST - allUnEd$sL2pitchST) < 0.8,
  allUnEd$sL1pitchST,
```

For further cases where the verb's original L1 was missing, the following line found other rises to the verb's peak that already started on the preceding subject, at the point originally determined as the subject's L2. This point was required to be realised late in the word, i.e. during the last third.[1]

```
ifelse(is.na(allUnEd$vL1pitchST)
   & (allUnEd$sL2relT > ((1/3)*2)), allUnEd$sL2pitchST,
```

In all other cases where the verb's original L1 was missing, it was substituted by the subject's L2 to reduce the number of missing values.

```
ifelse(is.na(allUnEd$vL1pitchST), allUnEd$sL2pitchST,
```

When the subject's L2 was located late in the word and had lower pitch than the originally measured L1 of the verb, the procedure assumed that the rise to the verb's peak already started at the end of the subject. The verb's 'real L1' was then determined to be the subject's L2.

```
ifelse((allUnEd$sL2relT > ((1/3)*2))
   & (allUnEd$vL1pitchST > allUnEd$sL2pitchST),
   allUnEd$sL2pitchST,
```

In other cases where the subject's L1 was missing, the code kept the verb's original L1 value.

```
ifelse(is.na(allUnEd$sL1pitchST), allUnEd$vL1pitchST,
```

When all values were available for these comparisons, the procedure identified 'rise to' contours on the subject as cases where the subject H was realised late in the word and had lower pitch than the verb's H. It was also required that no significant pitch fall intervened between the subject's and the verb's H. Then, the rise to the verb's peak was assumed to start with L1 measured on the subject.

```
ifelse((allUnEd$sHrelT > 1/3*2)
   & (allUnEd$sHpitchST < allUnEd$vHpitchST)
   & abs(allUnEd$sHpitchST - allUnEd$sL2pitchST) < 0.8
   & abs(allUnEd$sHpitchST - allUnEd$vL1pitchST) < 0.8,
   allUnEd$sL1pitchST,
```

In all remaining cases, the verb's 'real L1' was equated with its originally measured L1 value.

```
allUnEd$vL1pitchST))))))))))))
```

E.1.2 Minima following the peaks

For words with no or only an extremely short stretch of measurable f_0 (registered as `vHpitchST`), also the 'real' L2 of the verb was missing.

```
allUnEd$vL2Real <- as.numeric(ifelse(is.na(allUnEd$vHpitchST), NA,
```

[1] Note that in the absence of this line, the respective cases would have been detected by the following line, so that it is somewhat redundant. The author did however find it useful to get an overview of the data. For this she used an alternative version of the code that instead of outputting numerical values reported which line took effect for a particular word.

When no or only an extremely short stretch of measurable f_0 (registered as `oHpitchST`) was measurable for the object, L2 of the verb was concluded to be the originally measured value

```
ifelse(is.na(allUnEd$oHpitchST), allUnEd$vL2pitchST,
ifelse(is.na(allUnEd$oL1pitchST) & is.na(allUnEd$oL2pitchST),
   allUnEd$vL2pitchST,
```

Likewise, when verb and object were separated by a pause, the pitch fall from its peak was assumed not to have continued to the object.

```
ifelse(allUnEd$pause2=="yes", allUnEd$vL2pitchST,
```

In cases where neither L2 of the verb nor L1 of the object had been measured originally, the only possibility of detectable 'real' L2 value was a 'fall from' contour of the object, so that the end of the fall would be the object's L2. The code checked for the presence of a 'fall from' contour by examining whether the object's peak was early, and lower than the verb's peak. Note that the possibility of a significant rise interrupting the fall from the verb's H through the object's H can be practically excluded since none of the two possible intervening minima—L2 of the verb and L1 of the object—were detected.

```
ifelse(is.na(allUnEd$vL2pitchST) & is.na(allUnEd$oL1pitchST)
   & (allUnEd$oHrelT < 1/3)
   & (allUnEd$oHpitchST < allUnEd$vHpitchST),
   allUnEd$oL2pitchST,
```

When L1 of the object was missing, but L2 of the verb had been identified, a 'fall from' contour on the object could still be present when this L2 was not a real valley, but just a minor perturbation intervening. In addition to the criteria used in the previous line, the code therefore tested whether the fall from the verb's H to its L2 was not significantly larger than that to the object's H.

```
ifelse(is.na(allUnEd$oL1pitchST) & (allUnEd$oHrelT < 1/3)
   & (allUnEd$oHpitchST < allUnEd$vHpitchST)
   & ((allUnEd$vHpitchST - allUnEd$vL2pitchST)
   - (allUnEd$vHpitchST - allUnEd$oHpitchST)) < 0.8,
   allUnEd$oL2pitchST,
```

In all other cases, when L1 of the object had not been detected, a 'fall from' on the object could be excluded and the verb's 'real L2' was equated with the originally measured L2.

```
ifelse(is.na(allUnEd$oL1pitchST), allUnEd$vL2pitchST,
```

When the object's L1 had been measured (but the verb's L2 had not been), the procedure ascertained that it did not result from a real pitch fall when all the other criteria for a 'fall from' contour were met.

```
ifelse(is.na(allUnEd$vL2pitchST) & (allUnEd$oHrelT < 1/3)
   & (allUnEd$oHpitchST < allUnEd$vHpitchST)
   & abs(allUnEd$oHpitchST - allUnEd$oL1pitchST) < 0.8,
   allUnEd$oL2pitchST,
```

The fall from the verb's peak could also end with the object's L1. The code assumed this to be the case when it was realised early in the word, while L2 of the verb was missing.

```
ifelse(is.na(allUnEd$vL2pitchST) & (allUnEd$oL1relT < 1/3),
  allUnEd$oL1pitchST,
```

In all other cases where no L2 had originally been detected for the verb, the value of the object's L1 was substituted.

```
ifelse(is.na(allUnEd$vL2pitchST), allUnEd$oL1pitchST,
```

When the verb's L2 had been measured, the script required it to be higher than an early L1 on the object to conclude that the fall continued into the following word.

```
ifelse((allUnEd$oL1relT < 1/3)
  & (allUnEd$vL2pitchST > allUnEd$oL1pitchST),
  allUnEd$oL1pitchST,
ifelse(is.na(allUnEd$oL2pitchST), allUnEd$vL2pitchST,
```

For a 'fall from' contour on the object to be recognised when all values were available for comparison, the object's maximum needed to be realised early and with a lower f_0 than the verb's. Also, neither the rise from the object's L1 nor that from the verb's L2 to the object's H was allowed to be larger than 0.8 st.

```
ifelse((allUnEd$oHrelT < 1/3) &
  (allUnEd$oHpitchST < allUnEd$vHpitchST)
  & abs(allUnEd$oHpitchST - allUnEd$oL1pitchST) < 0.8
  & abs(allUnEd$oHpitchST - allUnEd$vL2pitchST) < 0.8,
  allUnEd$oL2pitchST,
```

In all remaining cases, the originally measured L2 of the verb was preserved.

```
allUnEd$vL2pitchST))))))))))))))
```

E.2 Evaluating the measured f_0 maxima

The peak was missing for words which were mostly or completely realised with non-modal voice quality.

```
allUnEd$vRealH <- as.numeric(ifelse(is.na(allUnEd$vHpitchST), NA,
  ifelse(is.na(allUnEd$vL1pitchST) & is.na(allUnEd$vL2pitchST),
    NA,
```

When the verb was preceded and followed by a pause or when no f_0 had been detected on the subject and the object, only the pitch range of the verb was relevant. When the range was smaller than 0.8 st, the pitch was considered to be flat and the f_0 maximum of was eliminated from the data set for the analysis of peak alignment. Otherwise it was retained.

```
ifelse(allUnEd$pause1 == "yes" & allUnEd$pause2 == "yes"
  & allUnEd$vrangeST >= 0.8,allUnEd$vHpitchST,
ifelse(allUnEd$pause1 == "yes" & allUnEd$pause2 == "yes", NA,
```

```
ifelse(is.na(allUnEd$sHpitchST) & is.na(allUnEd$oHpitchST)
  & allUnEd$vrangeST < 0.8, NA,
ifelse(is.na(allUnEd$sHpitchST) & is.na(allUnEd$oHpitchST),
  allUnEd$vHpitchST,
```

When there was a pause after the verb, its pitch contour was assumed to be independent from the following object, i.e. the possibility that it could carry a 'rise to' contour was excluded. The code thus checked whether its pitch contour was a fall from the subject's peak. When the verb's ('real') L1 was missing, the code verified that its f_0 maximum was early and lower than that of the subject. When L1 had been detected, it was additionally checked that it did not represent a noteworthy valley interrupting the fall.

```
ifelse(allUnEd$pause2 == "yes" & is.na(allUnEd$vL1Real)
  & allUnEd$vHrelT < 1/3
  & allUnEd$vHpitchST < allUnEd$sHpitchST, NA,
ifelse(allUnEd$pause2 == "yes" & allUnEd$vHrelT < 1/3
  & allUnEd$vHpitchST < allUnEd$sHpitchST
  & abs(allUnEd$vHpitchST - allUnEd$vL1Real) < 0.8, NA,
```

In the presence of a pause to its right, the verb could also have a maximum that was only an insignificant deviation from the surrounding minima (i.e. the 'real' minima determined above, which were either localised on the verb or on the neighbouring words). When the value for L1 was missing, the maximum was compared only to the following minimum (L2) and the other way around. When both were available, the code compared the maximum to each of them.

```
ifelse(allUnEd$pause2 == "yes" & is.na(allUnEd$vL1Real)
  & allUnEd$vrangeST < 0.8
  & (allUnEd$vHpitchST - allUnEd$vL2Real < 0.8), NA,
ifelse(allUnEd$pause2 == "yes" & is.na(allUnEd$vL2Real)
  & allUnEd$vrangeST < 0.8
  & (allUnEd$vHpitchST - allUnEd$vL1Real < 0.8), NA,
ifelse(allUnEd$pause2 == "yes" & allUnEd$vrangeST < 0.8
  & (allUnEd$vHpitchST - allUnEd$vL2Real < 0.8)
  & (allUnEd$vHpitchST - allUnEd$vL1Real < 0.8), NA,
```

After trying for all scenarios that would indicate that the measured maximum should be excluded from the analysis, the code admitted all other cases with a post-verbal pause.

```
ifelse(allUnEd$pause2 == "yes", allUnEd$vHpitchST,
```

When no pitch had been detected on the object at all, it was equally irrelevant in evaluating the status of the verb's pitch maximum. Like for the cases with an intervening pause, the possibility of the verb carrying a 'rise to' contour was excluded. Thus, its measured f_0 maximum could be unsuitable for the analysis of alignment either because it was part of a fall from the subject's peak or because the pitch was flat on a whole (the movements to and from the maximum being smaller than 0.8 st). The code checked for each of these possibilities and—when none of them was applicable—preserved the respective maximum in the same way as above for the presence of a pause.

```
ifelse(is.na(allUnEd$oHpitchST) & is.na(allUnEd$vL1Real)
  & allUnEd$vHrelT < 1/3
  & allUnEd$vHpitchST < allUnEd$sHpitchST, NA,
ifelse(is.na(allUnEd$oHpitchST) & allUnEd$vHrelT < 1/3
  & allUnEd$vHpitchST < allUnEd$sHpitchST
  & abs(allUnEd$vHpitchST - allUnEd$vL1Real) < 0.8, NA,
ifelse(is.na(allUnEd$oHpitchST) & is.na(allUnEd$vL1Real)
  & allUnEd$vrangeST < 0.8
  & (allUnEd$vHpitchST - allUnEd$vL2Real < 0.8), NA,
ifelse(is.na(allUnEd$oHpitchST) & is.na(allUnEd$vL2Real)
  & allUnEd$vrangeST < 0.8
  & (allUnEd$vHpitchST - allUnEd$vL1Real < 0.8), NA,
ifelse(is.na(allUnEd$oHpitchST) & allUnEd$vrangeST < 0.8
  & (allUnEd$vHpitchST - allUnEd$vL2Real < 0.8)
  & (allUnEd$vHpitchST - allUnEd$vL1Real < 0.8), NA,
ifelse(is.na(allUnEd$oHpitchST), allUnEd$vHpitchST,
```

With a pause preceding the verb, the situation mirrors that outlined above: The subject is considered irrelevant for the pitch contour of the verb, ruling out the possibility that its maximum was part of a 'fall from' contour. Conversely, the 'rise to' option is applicable, as is the possibility that the verb's pitch was flat, while all other measured f_0 maxima should be retained for the analysis of alignment.

The code therefore first tested for the presence of a 'rise to' contour on the verb in the absence of a measurable L2. It was taken to be present when the maximum was late in the word and had a lower f_0 than the object's maximum.

```
ifelse(allUnEd$pause1 == "yes" & is.na(allUnEd$vL2Real)
  & allUnEd$vHrelT > ((1/3)*2)
  & allUnEd$vHpitchST < allUnEd$oHpitchST, NA,
```

When L2 of the verb as measurable, the code ensured that it did not constitute a relevant intervening fall, in addition to the other test, before attesting that a 'rise to' contour was present.

```
ifelse(allUnEd$pause1 == "yes" & allUnEd$vHrelT > ((1/3)*2)
  & allUnEd$vHpitchST < allUnEd$oHpitchST
  & abs(allUnEd$vHpitchST - allUnEd$vL2Real) < 0.8, NA,
```

The possibility of a flat contour was first evaluated for cases without measurable L1. They only required a comparison between the maximum and the following minimum L2.

```
ifelse(allUnEd$pause1 == "yes" & is.na(allUnEd$vL1Real)
   & allUnEd$vrangeST < 0.8
   & (allUnEd$vHpitchST - allUnEd$vL2Real < 0.8), NA,
```

Conversely, H was only compared to L1 when L2 was missing.

```
ifelse(allUnEd$pause1 == "yes" & is.na(allUnEd$vL2Real)
   & allUnEd$vrangeST < 0.8
   & (allUnEd$vHpitchST - allUnEd$vL1Real < 0.8), NA,
```

A comparison to both minima was performed if possible.

```
ifelse(allUnEd$pause1 == "yes" & allUnEd$vrangeST < 0.8
   & (allUnEd$vHpitchST - allUnEd$vL2Real < 0.8)
   & (allUnEd$vHpitchST - allUnEd$vL1Real < 0.8), NA,
```

For all other verbs with a preceding pause, the maximum was included in the data set.

```
ifelse(allUnEd$pause1 == "yes", allUnEd$vHpitchST,
```

The situation was the same when no pitch could be detected for the subject. As above, the code tested for reasons to exclude the verb's H—because it was part of a 'rise to' contour or because pitch was too flat—and retained it when none of them applied.

```
ifelse(is.na(allUnEd$sHpitchST) & is.na(allUnEd$vL2Real)
   & allUnEd$vHrelT > ((1/3)*2)
   & allUnEd$vHpitchST < allUnEd$oHpitchST, NA,
ifelse(is.na(allUnEd$sHpitchST) & allUnEd$vHrelT > ((1/3)*2)
   & allUnEd$vHpitchST < allUnEd$oHpitchST
   & abs(allUnEd$vHpitchST - allUnEd$vL2Real) < 0.8, NA,
ifelse(is.na(allUnEd$sHpitchST) & is.na(allUnEd$vL1Real)
   & allUnEd$vrangeST < 0.8
   & (allUnEd$vHpitchST - allUnEd$vL2Real < 0.8), NA,
ifelse(is.na(allUnEd$sHpitchST) & is.na(allUnEd$vL2Real)
   & allUnEd$vrangeST < 0.8
   & (allUnEd$vHpitchST - allUnEd$vL1Real < 0.8), NA,
ifelse(is.na(allUnEd$sHpitchST) & allUnEd$vrangeST < 0.8
   & (allUnEd$vHpitchST - allUnEd$vL2Real < 0.8)
   & (allUnEd$vHpitchST - allUnEd$vL1Real < 0.8), NA,
ifelse(is.na(allUnEd$sHpitchST), allUnEd$vHpitchST,
```

Next, the code caught cases where the value for L1 was missing and checked whether they had a flat verb pitch or a 'fall from' contour.

```
ifelse(is.na(allUnEd$vL1Real) & allUnEd$vrangeST < 0.8
   & (allUnEd$vHpitchST - allUnEd$vL2Real < 0.8), NA,
```

```
ifelse(is.na(allUnEd$vL1Real) & allUnEd$vHrelT < 1/3
  & allUnEd$vHpitchST < allUnEd$sHpitchST, NA,
```

Then it tested for flat or 'rise to' contours for cases when no L2 had been measurable.

```
ifelse(is.na(allUnEd$vL2Real) & allUnEd$vrangeST < 0.8
  & (allUnEd$vHpitchST - allUnEd$vL1Real < 0.8), NA,
ifelse(is.na(allUnEd$vL2Real) & allUnEd$vHrelT > ((1/3)*2)
  & allUnEd$vHpitchST < allUnEd$oHpitchST, NA,
```

When all values were available for comparison, the procedure first filtered flat contours. Those were expected to have not only a small pitch range on the verb, but also at most a very small continuation of the fall to the following word or of the rise from the preceding one.

```
ifelse(allUnEd$vrangeST < 0.8
  & (allUnEd$vHpitchST - allUnEd$vL2Real < 0.8)
  & (allUnEd$vHpitchST - allUnEd$vL1Real < 0.8), NA,
```

It then tested whether H of the verb was part of a fall from the preceding subject's peak. This was assumed to be the case when the verb's pitch maximum was early, smaller than that of the subject and no noteworthy valley intervened.

```
ifelse(allUnEd$vHrelT < 1/3
  & allUnEd$vHpitchST < allUnEd$sHpitchST
  & abs(allUnEd$vHpitchST - allUnEd$vL1Real) < 0.8, NA,
```

The procedure considered a 'rise to' contour to be present on the verb when it's H was late and lower than that of the object without a significant fall intervening between the two maxima.

```
ifelse(allUnEd$vHrelT > ((1/3)*2)
  & allUnEd$vHpitchST < allUnEd$oHpitchST
  & abs(allUnEd$vHpitchST - allUnEd$vL2Real) < 0.8, NA,
```

In all other cases, the f_0 maximum of the verb remained in the data set for the analysis of peak alignment.

```
allUnEd$vHpitchST))))))))))))))))))))))))))))))))))))
```

Appendix F

Peak alignment for data sets edited with alternative criteria for perceptually relevant pitch movements

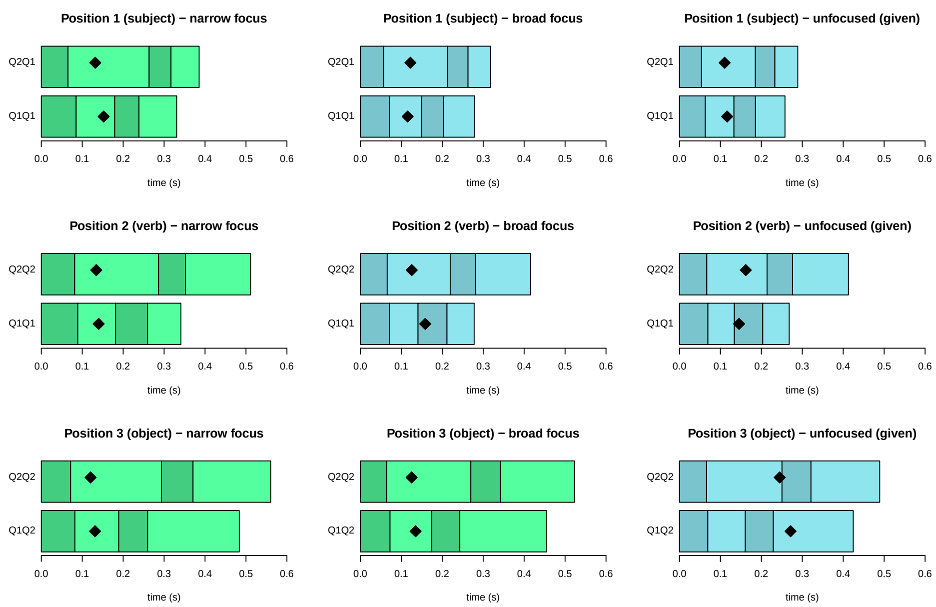

Figure F.1: Average segment durations and times of f0 turning points for all experimental conditions for the a differently edited data set with a criterion of 1 st instead of 0.8 st.

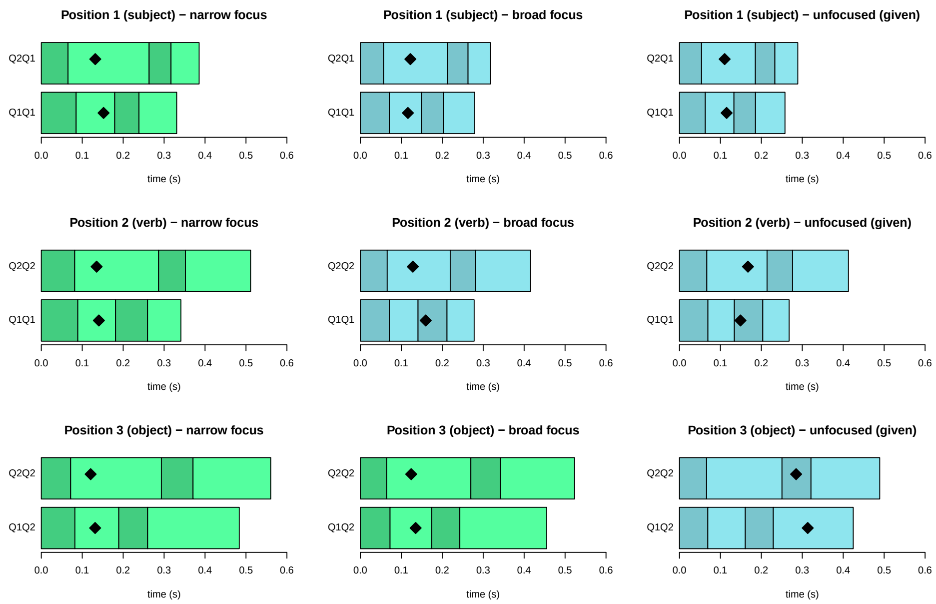

Figure F.2: Average segment durations and times of f0 turning points for all experimental conditions for the a differently edited data set with a criterion of 1.5 st instead of 0.8 st.

Appendix G

Random effects for models of peak alignment

Random by-subject adjustments for two linear mixed-effects models of peak alignment discussed in Chapter 3 are given below. Table G.1 lists the by-subject adjustments for the model of *H - Beginning of the nucleus* in Table 3.2, while G.2 gives the by-subject adjustments for the model of *H - End of the nucleus* in Table 3.3.

The function of these adjustments in statistical models is to account for individual (random) differences between participants in a study and specifically for the differences in how participants are influenced by the experimental variables. The random by-subject effects of models included a random interaction between the factors quantity and nuclear condition specified by the term `+ (1 + quantity * nuclear condition | subject)`. This means that the models estimated different effect sizes for the factors quantity and nuclear condition and that, additionally, the interplay between quantity and nuclear condition affected peak timing differently for the individual participants.

For example, the negative intercept value for subject s02 in Table G.1 indicates that this participant generally realised peaks earlier (i.e. closer to the beginning of the nucleus) than the model's estimated overall value based on the whole data set comprising all participants. The positive adjustment for the effect of quantity 2 for subject s02 in the same table means that the insignificant overall effect of quantity 2 in Table 3.2 was larger for this subject. Likewise, the negative adjustment for nuclear condition means that the overall significant negative effect appearing in Table 3.2 was also larger (more negative) for s02. Finally, the negative interaction adjustment for this participant indicates that the model estimated an additional adjustment for the peaks of words uttered by this subject that contained quantity 2 vowels *and* were in nuclear condition. That is, to arrive at the estimated value for the peak time of a quantity 2 word in nuclear condition spoken by s02, one needs to take the general model intercept and add the estimated effect for the factors quantity and nuclear condition, the intercept adjustment for s02, the adjustment of the factors quantity and nuclear condition for s02 *and additionally* adjust the value by −16.0814.

From this it is not possible to say whether there was a significant interaction for s02 or whether the interaction between the two factors was even significantly larger for this subject than for the data set on a whole. Instead, it merely means that the model fit improved significantly when taking these kinds of differences between subjects into account.

Table G.1: Random by-subject adjustments of the best linear mixed-effects model of the peak distance from the beginning of the first syllable nucleus.

	(Intercept)	Quantity 2	Nuclear condition	Interaction	Group
s02	−1.6708	20.7508	−21.5024	−16.0814	A
s03	−4.0136	−39.1681	34.7980	32.8436	B
s04	7.4100	−24.2084	29.4956	16.8643	B
s05	10.8631	36.9345	−27.0968	−33.4291	A
s06	−8.5733	−4.7894	−2.2728	6.8233	–
s07	15.4595	27.6720	−14.3889	−27.5883	A
s08	9.3478	18.9051	−10.8107	−18.4263	A
s09	−1.1176	−12.8540	11.5811	10.7092	B
s10	−14.9004	−25.3536	12.5888	25.5325	B
s11	−20.8076	−36.1614	18.3144	36.2622	B
s12	−11.0685	−2.9079	−6.1153	6.1791	–
s13	31.7920	44.5425	−17.5824	−46.8068	A
s14	17.4496	7.3284	6.9759	−11.9448	–
s15	−29.4714	−33.8065	9.0299	37.3804	B
s16	−17.4921	−28.6038	13.6522	29.0424	B
s17	−0.0843	8.8508	−8.6639	−7.0774	A
s18	18.6755	46.6016	−30.1757	−43.9046	A

To find out whether a significant interaction arose for a group of subjects, separate models were fit. Based on Tables G.1 and G.2, almost all subjects could readily be sorted into one of two groups. Group A showed positive adjustments for quantity and negative adjustments for nuclear condition and for the interaction between the two factors. For group B, the algebraic signs were inverse. Separate models were fit for both subject groups and both measures of alignment (cf. Tables G.3–G.6). None of these models showed an interaction between quantity and nuclear condition, neither in the fixed nor in the random effects. Instead, the emerging effects were similar to the models for the complete data set. Peaks were earlier in nuclear condition for group A, with a smaller distance from the beginning of the nucleus, but

Table G.2: Random by-subject adjustments of the best linear mixed-effects model of the peak distance from the end of the first syllable nucleus.

	(Intercept)	Quantity 2	Nuclear condition	Interaction	Group
s02	10.3625	36.5845	−20.1286	−33.4488	A
s03	−31.0448	−46.3342	27.0298	20.4545	B
s04	−1.4682	−15.1651	8.1013	17.3218	B
s05	10.0560	41.2722	−22.5675	−39.7326	A
s06	−1.0335	−1.3902	0.8197	0.4890	–
s07	8.4601	29.5514	−16.2667	−26.9088	A
s08	2.6703	9.4236	−5.1849	−8.6145	A
s09	−5.4710	−21.4272	11.7378	20.3220	B
s10	−4.9728	−25.8037	13.9967	26.4480	B
s11	−6.4026	−42.6282	22.9673	45.9083	B
s12	−4.5831	−2.5691	1.7442	−2.3644	–
s13	20.3679	49.4663	−27.7613	−37.4554	A
s14	15.9334	11.0474	−7.1764	5.5529	–
s15	−11.9402	−20.6064	11.8610	11.3787	B
s16	−15.1336	−16.0750	9.7518	1.7624	B
s17	15.7014	21.9887	−12.9105	−8.5229	A
s18	10.1540	37.3052	−20.4897	−34.6123	A

a larger distance from the end (cf. Tables G.3 and G.5, respectively). For group B, peaks only showed a larger distance from the end of the nucleus, but this effect was not significant as seen from the beginning of the nucleus (cf. Tables G.6 and G.4). While the two quantities did not differ regarding the distance of the peak from the beginning of the nucleus for either subject group, peaks of quantity 2 words—having longer vowel durations—showed a larger distance from the end of the nucleus than those of quantity 1 words.

Table G.3: Coefficients of the linear mixed-effects model of the peak distance from the beginning of the first syllable nucleus (in ms) for subject group A, random effects of subject and of nuclear condition by item.

	Estimate	Std. Error	t value
(Intercept)	138.6254	19.8649	6.9784
Quantity 2	3.9634	7.3163	0.5417
Nuclear condition	−87.2655	19.0089	−4.5908

Table G.4: Coefficients of the linear mixed-effects model of the peak distance from the beginning of the first syllable nucleus (in ms) for subject group B, random effects of nuclear condition by subject and of nuclear condition by item.

	Estimate	Std. Error	t value
(Intercept)	70.1330	8.8744	7.9029
Quantity 2	−7.8325	5.6550	−1.3850
Nuclear condition	−6.9087	10.7349	−0.6436

Table G.5: Coefficients of the linear mixed-effects model of the peak distance from the end of the first syllable nucleus (in ms) for subject group A, random effects of subject and of nuclear condition by item.

	Estimate	Std. Error	t value
(Intercept)	68.4500	16.8030	4.0737
Quantity 2	−95.2014	6.4493	−14.7615
Nuclear condition	−118.1478	18.4500	−6.4037

Table G.6: Coefficients of the linear mixed-effects model of the peak distance from the end of the first syllable nucleus (in ms) for subject group B, random effects of subject and of nuclear condition by item.

	Estimate	Std. Error	t value
(Intercept)	7.4367	9.6403	0.7714
Quantity 2	−113.9480	7.8789	−14.4625
Nuclear condition	−45.9312	8.1622	−5.6273

Appendix H

List of experimental sentences used in Production Experiment II

This Appendix presents the materials constituting the data of the production experiment analysed in Chapter 4. The following tables show all the sentences spoken by the participants together with the preceding prompt questions, in Finnish and with an English translation, in the second and first column from the right, respectively. The third column from the right gives the target word and its English translation separately. The English translations were intended to keep a balance between accuracy or closeness to the Finnish sentence and naturalness of the English equivalent. Note in particular that different renderings of passive verbs are possible; for example the passive past indicative of the verb *antaa* 'to give', *annettiin*, could be translated as '[it] was given', 'one gave', 'we gave' or 'you gave' (with a generic *you*).

The materials consisted of five sets, which each contained six item pairs with target words contrasting in first syllable structures: set 1 (CVVC.X vs. CV.X), set 2 (CVVC.X vs. CVV.X), set 3 (CVVC.X vs. CVC.X), set 4 (CVC.X vs. $\mathrm{CVC_g}$.X) and set 5 (CVV.X vs. CVC.X). Below, the first two columns from the left indicate the the first syllable contrast set ('Set') and item pair ('Pair') of the respective sentence. Each participant read all sentences from one of three item list; the third column from the left specifies this division of materials into lists.

Additional columns code, from left to right, focus condition ('Foc.'; B = broad or N = narrow), grammatical case (NOM = nominative, PRT = partitive or ESS = essive) and syllable structure of the target word ('Syll. Struc.').

Table H.1: List of items for Production Experiment II.

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
1	1	1	B	NOM	CVVC.CV	*saarna*	*Miksi kirkko oli täynnä?*	*Siellä kuultiin rovastin viimeisin saarna siveydestä.*
						'sermon'	'Why was the church full?'	'People were listening to the provost's last sermon about morality there.'
1	1	2	B	NOM	CV.CV	*sana*	*Miksi kirkko oli täynnä?*	*Siellä kuultiin rovastin viimeisin sana siveydestä.*
						'word'	'Why was the church full?'	'People were listening to the provost's last word about morality there.'
1	1	3	N	NOM	CVVC.CV	*saarna*	*Kuultiinko kirkossa rovastin viimeisin rukous siveydestä?*	*Siellä kuultiin rovastin viimeisin saarna siveydestä.*
						'sermon'	'Were people in the church listening to the provost's last prayer about morality?'	'People were listening to the provost's last sermon about morality there.'
1	1	1	N	NOM	CV.CV	*sana*	*Kuultiinko kirkossa rovastin viimeisin rukous siveydestä?*	*Siellä kuultiin rovastin viimeisin sana siveydestä.*
						'word'	'Were people in the church listening to the provost's last prayer about morality?'	'People were listening to the provost's last word about morality there.'
1	1	2	B	PRT	CVVC.CVV	*saarnaa*	*Mitä Petri teki sunnuntaina?*	*Hän kuunteli rovastin saarnaa siveydestä.*
						'sermon'	'What did Petri do on Sunday?'	'He listened to the provost's last sermon about morality.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
1	1	3	B	PRT	CV.CVV	*sanaa*	*Mitä Petri teki sunnuntaina?*	*Hän kuunteli rovastin sanaa siveydestä.*
						'word'	'What did Petri do on Sunday?'	'He listened to the provost's last word about morality.'
1	1	1	N	PRT	CVVC.CVV	'sermon'	*Kuunteliko Petri rovastin rukousta siveydestä?*	*Hän kuunteli rovastin saarnaa siveydestä.*
						'word'	'Did Petri hear the provost's last prayer about morality?'	'He listened to the provost's last sermon about morality.'
1	1	2	N	PRT	CV.CVV	*sanaa*	*Kuunteliko Petri rovastin rukousta siveydestä?*	*Hän kuunteli rovastin sanaa siveydestä.*
						'word'	'Did Petri hear the provost's last prayer about morality?'	'He listened to the provost's last word about morality.'
1	1	3	B	ESS	CVVC.CV.CV	*saarnana*	*Miksi siivooja ei heittänyt paperia roskiin?*	*Kotiapulainen piti tekstiä rovastin saarnana siveydestä.*
						'as a sermon'	'Why did the cleaner not throw away the paper?'	'The domestic thought the text was the provost's last sermon about morality.'
1	1	1	B	ESS	CV.CV.CV	*sanana*	*Miksi siivooja ei heittänyt paperia roskiin?*	*Kotiapulainen piti tekstiä rovastin sanana siveydestä.*
						'as a word'	'Why did the cleaner not throw away the paper?'	'The domestic thought the text was the provost's last word about morality.'
1	1	2	N	ESS	CVVC.CV.CV	*saarnana*	*Pitikö kotiapulainen tekstiä rovastin rukouksena siveydestä?*	*Kotiapulainen piti tekstiä rovastin saarnana siveydestä.*
						'as a sermon'	'Did the domestic think the text was the provost's prayer about morality?'	'The domestic thought the text was the provost's last sermon about morality.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
1	1	3	N	ESS	CV.CV.CV	*sanana* 'as a word'	*Pitikö kotiapulainen tekstiä rovastin rukouksena siveydestä?* 'Did the domestic think the text was the provost's prayer about morality?'	*Kotiapulainen piti tekstiä rovastin sanana siveydestä.* 'The domestic thought the text was the provost's last word about morality.'
1	2	2	B	NOM	CVVC.CV	*naarmu* 'scar'	*Miksi äiti torui poikaa?* 'Why did the mother scold the boy?'	*Pojalla oli kädessä naarmu tällä kertaa.* 'The boy had a scratch in his hand this time.'
1	2	3	B	NOM	CV.CV	*namu* 'candy'	*Miksi äiti torui poikaa?* 'Why did the mother scold the boy?'	*Pojalla oli kädessä namu tällä kertaa.* 'The boy had a candy in his hand this time.'
1	2	1	N	NOM	CVVC.CV	*naarmu* 'scar'	*Oliko pojalla kädessä multaa tällä kertaa?* 'Did the boy have soil in his hand this time?'	*Pojalla oli kädessä naarmu tällä kertaa.* 'The boy had a scratch in his hand this time.'
1	2	2	N	NOM	CV.CV	*namu* 'candy'	*Oliko pojalla kädessä lelu tällä kertaa?* 'Did the boy have a toy in his hand this time?'	*Pojalla oli kädessä namu tällä kertaa.* 'The boy had a candy in his hand this time.'
1	2	3	B	PRT	CVVC.CVV	*naarmua* 'scar'	*Miksi äiti torui poikaa?* 'Why did the mother scold the boy?'	*Poika piilotteli kädessä naarmua täysin pokkana.* 'The boy hid a scratch in his hand with a completely straight face.'
1	2	1	B	PRT	CV.CVV	*namua*	*Miksi äiti torui poikaa?*	*Poika piilotteli kädessä namua täysin pokkana.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						‘candy’	‘Why did the mother scold the boy?’	‘The boy hid a candy in his hand with a completely straight face.’
1	2	2	N	PRT	CVVC.CVV	*naarmua*	*Piilotteliko poika kädessä luomea täysin pokkana?*	*Poika piilotteli kädessä naarmua täysin pokkana.*
						‘scar’	‘Did the boy hide a mole on his hand with a completely straight face?’	‘The boy hid a scratch in his hand with a completely straight face.’
1	2	3	N	PRT	CV.CVV	*namua*	*Piilotteliko poika kädessä lelua täysin pokkana?*	*Poika piilotteli kädessä namua täysin pokkana.*
						‘candy’	‘Did the boy hide a toy on his hand with a completely straight face?’	‘The boy hid a candy in his hand with a completely straight face.’
1	2	1	B	ESS	CVVC.CV.CV	*naarmuna*	*Kuinka poika harhautti pikku veljensä?*	*Poika esitteli multaa kädessä naarmuna täysin pokkana.*
						‘as a scar’	‘How did the boy deceive his little brother?’	‘The boy presented soil in his hand as a scar with a completely straight face.’
1	2	2	B	ESS	CV.CV.CV	*namuna*	*Kuinka poika harhautti pikku veljensä?*	*Poika esitteli kiveä kädessä namuna täysin pokkana.*
						‘as candy’	‘How did the boy deceive his little brother?’	‘The boy presented a stone in his hand as candy with a completely straight face.’
1	2	3	N	ESS	CVVC.CV.CV	*naarmuna*	*Esittelikö poika multaa kädessä luomena täysin pokkana?*	*Poika esitteli multaa kädessä naarmuna täysin pokkana.*
						‘as a scar’	‘Did the boy present soil in his hand as a mole with a completely straight face?’	‘The boy presented soil in his hand as a scar with a completely straight face.’
1	2	1	N	ESS	CV.CV.CV	*namuna*	*Esittelikö poika kiveä kädessä leluna täysin pokkana?*	*Poika esitteli kiveä kädessä namuna täysin pokkana.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'as candy'	'Did the boy present a stone in his hand as a toy with a completely straight face?'	'The boy presented a stone in his hand as candy with a completely straight face.'
1	3	3	B	NOM	CVVC.CV	*vaarna*	*Kuinka tiesit, että rannalle oli leiriy-dytty?*	*Partiolaisilta jäi hiekalle vaarna selvästi näkyville.*
						'tent peg'	'How did you know that someone had camped on the beach?'	'The scouts left a tent peg in the sand in plain sight.'
1	3	1	B	NOM	CV.CV	*vana*	*Kuinka tiesit, että rannalle oli leiriy-dytty?*	*Partiolaisilta jäi hiekalle vana selvästi näkyville.*
						'trail'	'How did you know that someone had camped on the beach?'	'The scouts left a trail in the sand in plain sight.'
1	3	2	N	NOM	CVVC.CV	*vaarna*	*Jäikö partiolaisilta hiekalle teltta selvästi näkyville?*	*Partiolaisilta jäi hiekalle vaarna selvästi näkyville.*
						'tent peg'	'Did the scouts leave a tent in the sand in plain sight?'	'The scouts left a tent peg in the sand in plain sight.'
1	3	3	N	NOM	CV.CV	*vana*	*Jäikö partiolaisilta hiekalle nuotio selvästi näkyville?*	*Partiolaisilta jäi hiekalle vana selvästi näkyville.*
						'trail'	'Did the scouts leave a camp fire in the sand in plain sight?'	'The scouts left a trail in the sand in plain sight.'
1	3	1	B	PRT	CVVC.CVV	*vaarnaa*	*Kuinka tiesit, että rannalle oli leiriy-dytty?*	*Partiolaiset näyttivät minulle vaarnaa selvänä todisteena.*
						'tent peg'	'How did you know that someone had camped on the beach?'	'The scouts showed me a tent peg as a clear piece of evidence.'
1	3	2	B	PRT	CV.CVV	*vanaa*	*Kuinka tiesit, että rannalle oli leiriy-dytty?*	*Partiolaiset näyttivät minulle vanaa selvänä todisteena.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'trail'	'How did you know that someone had camped on the beach?'	'The scouts showed me a trail as a clear piece of evidence.'
1	3	3	N	PRT	CVVC.CVV	*vaarnaa*	*Näyttivätkö partiolaiset sinulle telttaa selvänä todisteena?*	*Partiolaiset näyttivät minulle vaarnaa selvänä todisteena.*
						'tent peg'	'Did the scouts show you a a tent as a clear piece of evidence?'	'The scouts showed me a tent peg as a clear piece of evidence.'
1	3	1	N	PRT	CV.CVV	*vanaa*	*Näyttivätkö partiolaiset sinulle nuotiota selvänä todisteena?*	*Partiolaiset näyttivät minulle vanaa selvänä todisteena.*
						'trail'	'Did the scouts show you a camp fire as a clear piece of evidence.'	'The scouts showed me a trail as a clear piece of evidence.'
1	3	2	B	ESS	CVVC.CV.CV	*vaarnana*	*Mitä partiolaiset sanoivat?*	*Ne esittelivät kepin johtajalle vaarnana selkeästi perustellen.*
						'as a tent peg'	'What did the scouts say?'	'They presented the stick to the leader as a tent peg with a clear justification.'
1	3	3	B	ESS	CV.CV.CV	*vanana*	*Mitä partiolaiset sanoivat?*	*Ne esittelivät jäljen johtajalle vanana selkeästi perustellen.*
						'as a trail'	'What did the scouts say?'	'They presented the track to the leader as a trail with a clear justification.'
1	3	1	N	ESS	CVVC.CV.CV	*vaarnana*	*Esittelivätkö partiolaiset kepin nuolena selkeästi perustellen?*	*Ne esittelivät kepin johtajalle vaarnana selkeästi perustellen.*
						'as a tent peg'	'Did the scouts present the stick to the leader as a camp fire with a clear justification?'	'They presented the stick to the leader as a tent peg with a clear justification.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
1	3	2	N	ESS	CV.CV.CV	*vanana*	*Esittelivätkö partiolaiset jäljen poterona selkeästi perustellen?*	*Ne esittelivät jäljen johtajalle vanana selkeästi perustellen.*
						'as a trail'	'Did the scouts present the track to the leader as a camp fire with a clear justification?'	'They presented the track to the leader as a trail with a clear justification.'
1	4	1	B	NOM	CVVC.CV	*kaarna*	*Miksi lapset nauroivat?*	*Koiralla oli suussa kaarna pureskeltavana.*
						'bark'	'Why did the children laugh?'	'The dog had a bark in its mouth while chewing.'
1	4	3	B	NOM	CV.CV	*kana*	*Miksi lapset nauroivat?*	*Koiralla oli suussa kana pureskeltavana.*
						'chicken'	'Why did the children laugh?'	'The dog had a chicken in its mouth while chewing.'
1	4	2	N	NOM	CVVC.CV	*kaarna*	*Oliko koiralla suussa risu pureskeltavana?*	*Koiralla oli suussa kaarna pureskeltavana.*
						'bark'	'Did the dog have a twig in its mouth while chewing?'	'The dog had a bark in its mouth while chewing.'
1	4	1	N	NOM	CV.CV	*kana*	*Oliko koiralla suussa pulu pureskeltavana?*	*Koiralla oli suussa kana pureskeltavana.*
						'chicken'	'Did the dog have a pigeon in its mouth while chewing?'	'The dog had a chicken in its mouth while chewing.'
1	4	3	B	PRT	CVVC.CVV	*kaarnaa*	*Miksi lapset nauroivat?*	*Koira kantoi suussa kaarnaa pureskeltavana.*
						'bark'	'Why did the children laugh?'	'The dog carried a bark in its mouth while chewing.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
1	4	2	B	PRT	CV.CVV	*kanaa*	*Miksi lapset nauroivat?*	*Koira kantoi suussa kanaa pureskeltavana.*
						'chicken'	'Why did the children laugh?'	'The dog carried a chicken in its mouth while chewing.'
1	4	1	N	PRT	CVVC.CVV	*kaarnaa*	*Kantoiko koira suussa risua pureskeltavana?*	*Koira kantoi suussa kaarnaa pureskeltavana.*
						'bark'	'Did the dog carry a twig in its mouth while chewing?'	'The dog carried a bark in its mouth while chewing.'
1	4	3	N	PRT	CV.CVV	*kanaa*	*Kantoiko koira suussa pulua pureskeltavana?*	*Koira kantoi suussa kanaa pureskeltavana.*
						'chicken'	'Did the dog carry a pigeon in its mouth while chewing?'	'The dog carried a chicken in its mouth while chewing.'
1	4	2	B	ESS	CVVC.CV.CV	*kaarnana*	*Miksi poika oli hämmästynyt?*	*Poika piti puunkuorta liiterissä kaarnana puolivahingossa.*
						'as bark'	'Why was the boy surprised?'	'The boy (mis)took the rind in the shed for bark almost by accident.'
1	4	1	B	ESS	CV.CV.CV	*kanana*	*Miksi isä oli yllättynyt?*	*Isä söi tofua ravintolassa kanana puolivahingossa.*
						'as a chicken'	'Why was father surprised?'	'father ate tofu in the restaurant thinking it was chicken [lit. 'as chicken'] almost by accident.'
						'as bark'	'Did the boy (mis)take the rind in the shed for birchbark almost by accident?'	'The boy (mis)took the rind in the shed for bark almost by accident.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
1	4	2	N	ESS	CV.CV.CV	*kanana*	*Söikö isä ravintolassa tofua possuna puolivahingossa?*	*Isä söi tofua ravintolassa kanana puolivahingossa.*
						'as a chicken'	'Did father eat tofu in the restaurant thinking it was pork [lit. 'as pork'] almost by accident.'	'father ate tofu in the restaurant thinking it was chicken [lit. 'as chicken'] almost by accident.'
1	5	2	B	NOM	CVVC.CV	*Kaarle*	*Mitä siskosi kertoi Kalesta?*	*Tutuille hän on työpaikalla Kaarle kutsumanimeltään.*
						'Kaarle' (name)	'What did your sister tell about Kale?'	'To aquaintances, his nickname is Kaarle [lit. 'he is K. as a his nickname'] at the workplace.'
1	5	1	B	NOM	CV.CV	*Kale*	*Mitä siskosi kertoi Kaarlesta?*	*Tutuille hän on työpaikalla Kale kutsumanimeltään.*
						'Kale' (name)	'What did your sister tell about Kaarle?'	'To aquaintances, his nickname is Kale [lit. 'he is K. as a his nickname'] at the workplace.'
1	5	3	N	NOM	CVVC.CV	*Kaarle*	*Onko hän tutuille työpaikalla Pentti kutsumanimeltään?*	*Tutuille hän on työpaikalla Kaarle kutsumanimeltään.*
						'Kaarle' (name)	'Is his his nickname Pentti [lit. 'he is P. as a his nickname'] to aquaintances at the workplace?'	'To aquaintances, his nickname is Kaarle [lit. 'he is K. as a his nickname'] at the workplace.'
1	5	2	N	NOM	CV.CV	*Kale*	*Onko hän tutuille työpaikalla Pentti kutsumanimeltään?*	*Tutuille hän on työpaikalla Kale kutsumanimeltään.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'Kale' (name)	'Is his his nickname Pentti [lit. 'he is P. as a his nickname'] to aquaintances at the workplace?'	'To aquaintances, his nickname is Kale [lit. 'he is K. as a his nickname'] at the workplace.'
1	5	1	B	PRT	CVVC.CVV	*Kaarlea*	*Mitä ne tekivät metsässä?*	*Siellä haettiin porukalla Kaarlea kuumeisesti.*
						'Kaarle' (name)	'What did they do in the forest?'	'There the group group was looking for Kaarle feverishly.'
1	5	3	B	PRT	CV.CVV	*Kalea*	*Mitä ne tekivät metsässä?*	*Siellä haettiin porukalla Kalea kuumeisesti.*
						'Kale' (name)	'What did they do in the forest?'	'There the group group was looking for Kale feverishly.'
1	5	2	N	PRT	CVVC.CVV	*Kaarlea*	*Haettiinko siellä porukalla Penttiä kuumeisesti?*	*Siellä haettiin porukalla Kaarlea kuumeisesti.*
						'Kaarle' (name)	'Was the group group was looking for Pentti there feverishly?'	'There the group group was looking for Kaarle feverishly.'
1	5	1	N	PRT	CV.CVV	*Kalea*	*Haettiinko siellä porukalla Penttiä kuumeisesti?*	*Siellä haettiin porukalla Kalea kuumeisesti.*
						'Kale' (name)	'Was the group group was looking for Pentti there feverishly?'	'There the group group was looking for Kale feverishly.'
1	5	3	B	ESS	CVVC.CV.CV	*Kaarlena*	*Miksi Pentti oli pettynyt?*	*Häntä pidettiin porukalla Kaarlena kutsumanimeltään.*
						'as Kaarle' (name)	'Why was Pentti disappointed?'	'The group thought his nickname was Kaarle.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
1	5	2	B	ESS	CV.CV.CV	*Kalena*	*Miksi Pentti oli pettynyt?*	*Häntä pidettiin porukalla Kalena kutsumanimeltään.*
						'as Kale' (name)	'Why was Pentti disappointed?'	'The group thought his nickname was Kale.'
1	5	1	N	ESS	CVVC.CV.CV	*Kaarlena*	*Pidettiinkö häntä porukalla Penttinä kutsumanimeltään?*	*Häntä pidettiin porukalla Kaarlena kutsumanimeltään.*
						'as Kaarle' (name)	'Did the group think his nickname was Pentti?'	'The group thought his nickname was Kaarle.'
1	5	3	N	ESS	CV.CV.CV	*Kalena*	*Pidettiinkö häntä porukalla Penttinä kutsumanimeltään?*	*Häntä pidettiin porukalla Kalena kutsumanimeltään.*
						'as Kale' (name)	'Did the group think his nickname was Pentti?'	'The group thought his nickname was Kale.'
1	6	3	B	NOM	CVVC.CV	*jaarli*	*Miksi luulet isoäidin olevan hämmentynyt?*	*Mikko oli hänelle eilen jaarli kohtalaisen varmasti.*
						'earl'	'Why do you think grandmother is confused?'	'Mikko was an earl to her yesterday fairly certainly.'
1	6	2	B	NOM	CV.CV	*Jali*	*Miksi luulet isoäidin olevan hämmentynyt?*	*Mikko oli hänelle eilen Jali kohtalaisen varmasti.*
						'Jali' (name)	'Why do you think grandmother is confused?'	'Mikko was Jali to her yesterday fairly certainly.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
1	6	1	N	NOM	CVVC.CV	*jaarli*	*Oliko Mikko isoäidille eilen prinssi kohtalaisen varmasti?*	*Mikko oli hänelle eilen jaarli kohtalaisen varmasti.*
						'earl'	'Was Mikko an prince to grandmother yesterday fairly certainly.'	'Mikko was an earl to her yesterday fairly certainly.'
1	6	3	N	NOM	CV.CV	*Jali*	*Oliko Mikko isoäidille eilen Ville kohtalaisen varmasti?*	*Mikko oli hänelle eilen Jali kohtalaisen varmasti.*
						'Jali' (name)	'Was Mikko Ville to grandmother yesterday fairly certainly.'	'Mikko was Jali to her yesterday fairly certainly.'
1	6	2	B	PRT	CVVC.CVV	*jaarlia*	*Miksi Marko on hermostunut?*	*Elisa tapaili ennen jaarlia kohtalaisen useasti.*
						'earl'	'Why was Marko nervous?'	'Elisa dated an earl before fairly certainly.'
1	6	1	B	PRT	CV.CVV	*Jalia*	*Miksi Marko on hermostunut?*	*Elisa tapaili ennen Jalia kohtalaisen useasti.*
						'Jali' (name)	'Why was Marko nervous?'	'Elisa dated Jali before fairly certainly.'
1	6	3	N	PRT	CVVC.CVV	*jaarlia*	*Tapailiko Elisa ennen prinssiä kohtalaisen useasti?*	*Elisa tapaili ennen jaarlia kohtalaisen useasti.*
						'earl'	'Did Elisa date a prince before fairly certainly?'	'Elisa dated an earl before fairly certainly.'
1	6	2	N	PRT	CV.CVV	*Jalia*	*Tapailiko Elisa ennen Villeä kohtalaisen useasti?*	*Elisa tapaili ennen Jalia kohtalaisen useasti.*
						'as Jali' (name)	'Did Elisa date Ville before fairly certainly?'	'Elisa dated Jali before fairly certainly.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
1	6	1	B	ESS	CVVC.CV.CV	*jaarlina*	*Miksi Elisa oli vihainen?*	*Mikko esiintyi eilen jaarlina kohtalaisen julkeasti.*
						'as an earl'	'Why was Elisa angry?'	'Mikko presented himself as an earl yesterday fairly certainly.'
1	6	3	B	ESS	CV.CV.CV	*Jalina*	*Miksi Elisa oli vihainen?*	*Mikko esiintyi eilen Jalina kohtalaisen julkeasti.*
						'as Jali' (name)	'Why was Elisa angry?'	'Mikko presented himself as Jali yesterday fairly certainly.'
1	6	2	N	ESS	CVVC.CV.CV	*jaarlina*	*Esiintyikö Mikko eilen prinssinä kohtalaisen julkeasti?*	*Mikko esiintyi eilen jaarlina kohtalaisen julkeasti.*
						'as an earl'	'Did Mikko present himself as a prince yesterday fairly certainly?'	'Mikko presented himself as an earl yesterday fairly certainly.'
1	6	1	N	ESS	CV.CV.CV	*Jalina*	*Esiintyikö Mikko eilen prinssinä kohtalaisen julkeasti?*	*Mikko esiintyi eilen Jalina kohtalaisen julkeasti.*
						'as Jali' (name)	'Did Mikko present himself as a prince yesterday fairly certainly?'	'Mikko presented himself as Jali yesterday fairly certainly.'
2	1	1	B	NOM	CVVC.CV	*saarna*	*Mihin Miia keskittyi?*	*Taustalla kuului papin mahtava saarna koko aamun.*
						'sermon'	'What did Miia concentrate on?'	'In the background the priest's magnificent sermon was audible the whole morning.'
2	1	2	B	NOM	CVV.CV	*Saana*	*Mihin Miia keskittyi?*	*Taustalla kuulsi Lapin mahtava Saana koko aamun.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'Saana' (a fell in Lapland)	'What did Miia concentrate on?'	'In the background the magnificent Saana of Lapland [a fell] was dimly visible the whole morning.'
2	1	3	N	NOM	CVVC.CV	*saarna*	*Kuuluiko taustalla rovastin mahtava rukous koko aamun?*[1]	*Taustalla kuului papin mahtava saarna koko aamun.*
						'sermon'	'Was the provost's magnificent prayer audible in the background the whole morning?'	'In the background the priest's magnificent sermon was audible the whole morning.'
2	1	1	N	NOM	CVV.CV	*Saana*	*Kuulsiko taustalla Lapin mahtava Ylläs koko aamun?*	*Taustalla kuulsi Lapin mahtava Saana koko aamun.*
						'Saana' (a fell in Lapland)	'Was the magnificent Ylläs of Lapland [a fell] dimly visible in the background the whole morning?'	'In the background the magnificent Saana of Lapland [a fell] was dimly visible the whole morning.'
2	1	2	B	PRT	CVVC.CVV	*saarnaa*	*Mitä eilen tehtiin?*	*Eilen kuunneltiin papin mahtavaa saarnaa koko aamun.*
						'sermon'	'What did people do yesterday?'	'Yesterday people listened to the priest's magnificent sermon the whole morning.'
2	1	3	B	PRT	CVV.CVV	*Saanaa*	*Mitä eilen tehtiin?*	*Eilen katseltiin Lapin mahtavaa Saanaa koko aamun.*

Continued on next page

[1] This prompt should have included the word *papin* 'priest's' instead of *rovastin* 'provost's'. Nevertheless, almost all participants produced the intended focus structure, so that materials were included into the analyses in spite of this mistake.

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'Saana' (a fell in Lapland)	'What did people do yesterday?'	'Yesterday people looked at the magnificent Saana of Lapland [a fell] the whole morning.'
2	1	1	N	PRT	CVVC.CVV	*saarnaa*	*Kuunneltiinko eilen papin mahtavaa rukousta koko aamun?*	*Eilen kuunneltiin papin mahtavaa saarnaa koko aamun.*
						'sermon'	'Did people listen to the priest's magnificent prayer the whole morning yesterday?'	'Yesterday people listened to the priest's magnificent sermon the whole morning.'
2	1	2	N	PRT	CVV.CVV	*Saanaa*	*Katseltiinko eilen lapin mahtavaa Yllästä koko aamun?*	*Eilen katseltiin Lapin mahtavaa Saanaa koko aamun.*
						'Saana' (a fell in Lapland)	'Did people look at the magnificent Ylläs of Lapland [a fell] the whole morning yesterday?'	'Yesterday people looked at the magnificent Saana of Lapland [a fell] the whole morning.'
2	1	3	B	ESS	CVVC.CV.CV	*saarnana*	*Mikä se ääni oli?*	*Eilen pidettiin sitä papin mahtavana saarnana koko aamun.*
						'as a sermon'	'What was this sound?'	'Yesterday people thought it was the priest's magnificent sermon the whole morning.'
2	1	1	B	ESS	CVV.CV.CV	*Saanana*	*Mitä siellä näkyi?*	*Eilen pidettiin sitä Lapin mahtavana Saanana koko aamun.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'as Saana' (a fell in Lapland)	'What was visible there?'	'Yesterday people thought it was the magnificent Saana of Lapland [a fell] the whole morning.'
2	1	2	N	ESS	CVVC.CV.CV	*saarnana*	*Pidettiinkö ääntä eilen papin mahtavana rukouksena koko aamun?*	*Eilen pidettiin sitä papin mahtavana saarnana koko aamun.*
						'as a sermon'	'Did people think the sound was the priest's magnificent prayer yesterday the whole morning?'	'Yesterday people thought it was the priest's magnificent sermon the whole morning.'
2	1	3	N	ESS	CVV.CV.CV	*Saanana*	*Pidettiinkö näkymää eilen Lapin mahtavana Ylläksenä koko aamun?*	*Eilen pidettiin sitä Lapin mahtavana Saanana koko aamun.*
						'as Saana' (a fell in Lapland)	'Did people think the outlook was the magnificent Ylläs of Lapland [a fell] the whole morning yesterday?'	'Yesterday people thought it was the magnificent Saana of Lapland [a fell] the whole morning.'
2	2	2	B	NOM	CVVC.CV	*kuorma*	*Miksi rekka ei ajanut naapurin pihalle?*	*Siellä oli hiekkaa ennestään vanha kuorma pitkän aikaa.*
						'truck'	'Why did the truck not drive on the neighbour's yard?'	'There was already an old load of sand for a long time.'
2	2	3	B	NOM	CVV.CV	*kuoma*	*Miksi perheesi tykkää naapurin isännästä?*	*Hän oli Penalle ennestään vanha kuoma pitkän aikaa.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'chum'	'Why does your family like the neighbour [lit. 'the male head of the neighbour household']?'	'He was already an old buddy of Pena's for a long time.'
2	2	1	N	NOM	CVVC.CV	*kuorma*	*Oliko naapurin pihalla hiekkaa ennestään vanha keko pitkän aikaa?*	*Siellä oli hiekkaa ennestään vanha kuorma pitkän aikaa.*
						'truck'	'Was there already an old heap of sand on the neighbour's yard for a long time?'	'There was already an old load of sand for a long time.'
2	2	2	N	NOM	CVV.CV	*kuoma*	*Oliko naapurin isäntä Penalle ennestään vanha vihamies pitkän aikaa?*	*Hän oli Penalle ennestään vanha kuoma pitkän aikaa.*
						'chum'	'Was the neighbour [lit. 'the male head of the neighbour household'] already an old enemy of Pena's for a long time?'	'He was already an old buddy of Pena's for a long time.'
2	2	3	B	PRT	CVVC.CVV	*kuormaa*	*Miksi rekka ajoi naapurin pihalle?*	*Siellä oleva hiekka oli osa vanhaa kuormaa pitkän aikaa.*
						'truck'	'Why did the truck drive on the neighbour's yard?'	'The sand that was there was part of an old load for a long time.'
2	2	1	B	PRT	CVV.CVV	*kuomaa*	*Miksi perheesi tykkää naapurin isännästä?*	*Hän oli Penalle ennestään vanhaa kuomaa pitkän aikaa.*
						'chum'	'Why does your family like the neighbour [lit. 'the male head of the neighbour household']?'	'He was already an old buddy of Pena's for a long time.'
2	2	2	N	PRT	CVVC.CVV	*kuormaa*	*Oliko naapurin pihalla oleva hiekka osa vanhaa kekoa pitkän aikaa?*	*Siellä oleva hiekka oli osa vanhaa kuormaa pitkän aikaa.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'truck'	'Was the sand that was there was part of an old heap for a long time?'	'The sand that was there was part of an old load for a long time.'
2	2	3	N	PRT	CVV.CVV	*kuomaa*	*Oliko hän Penalle ennestään vanhaa vihamiestä pitkän aikaa?*	*Hän oli Penalle ennestään vanhaa kuomaa pitkän aikaa.*
						'chum'	'Was he already an old enemy of Pena's for a long time?'	'He was already an old buddy of Pena's for a long time.'
2	2	1	B	ESS	CVVC.CV.CV	*kuormana*	*Miksi naapurin piha on niin täynnä?*	*Sinne toimitettiin hiekkaa suurena kuormana pitkän aikaa.*
						'as a truck'	'Why is the neighbour's yard this full?'	'A big load of sand [lit. 'sand as a big load'] was delivered there a long time ago.'
2	2	2	B	ESS	CVV.CV.CV	*kuomana*	*Miksi kaverit tykkäävät naapurin isännästä?*	*Hänet tunnettiin yleisesti vanhana kuomana pitkän aikaa.*
						'as a chum'	'Why do the guys like the neighbour [lit. 'the male head of the neighbour household']?'	'He was generally known as an old buddy of Pena's for a long time.'
2	2	3	N	ESS	CVVC.CV.CV	*kuormana*	*Toimitettiinko sinne hiekkaa suurena kekona pitkän aikaa?*	*Sinne toimitettiin hiekkaa suurena kuormana pitkän aikaa.*
						'as a truck'	'Was big heap of sand [lit. 'sand as a big load'] delivered there a long time ago?'	'A big load of sand [lit. 'sand as a big load'] was delivered there a long time ago.'
2	2	1	N	ESS	CVV.CV.CV	*kuomana*	*Tunnettiinko hänet yleisesti vanhana vihamiehenä pitkän aikaa?*	*Hänet tunnettiin yleisesti vanhana kuomana pitkän aikaa.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'as a chum'	'Was he generally known as an old enemy of Pena's for a long time.'	'He was generally known as an old buddy of Pena's for a long time.'
2	3	3	B	NOM	CVVC.CV	*luonto*	*Miksi Kaisa osti postikortin?*	*Kortissa näkyi rannikolla oleva luonto täydellisesti.*
						'nature'	'Why did Kaisa buy the postcard?'	'On the [post]card, nature at the shore was visible perfectly.'
2	3	1	B	NOM	CVV.CV	*luoto*	*Miksi Kaisa osti postikortin?*	*Kortissa näkyi rannikolla oleva luoto täydellisesti.*
						'islet'	'Why did Kaisa buy the postcard?'	'On the [post]card, the islet at the shore was visible perfectly.'
2	3	2	N	NOM	CVVC.CV	*luonto*	*Näkyikö kortissa rannikolla oleva kaupunki täydellisesti?*	*Kortissa näkyi rannikolla oleva luonto täydellisesti.*
						'nature'	'Was the city at the shore visible perfectly on the [post]card?'	'On the [post]card, nature at the shore was visible perfectly.'
2	3	3	N	NOM	CVV.CV	*luoto*	*Näkyikö kortissa rannikolla oleva kappeli täydellisesti?*	*Kortissa näkyi rannikolla oleva luoto täydellisesti.*
						'islet'	'Was the chapel at the shore visible perfectly on the [post]card?'	'On the [post]card, the islet at the shore was visible perfectly.'
2	3	1	B	PRT	CVVC.CVV	*luontoa*	*Miksi Kaisa osti postikortin?*	*Kortissa kuvattiin rannikolla olevaa luontoa täydellisesti.*
						'nature'	'Why did Kaisa buy the postcard?'	'The [post]card pictured nature at the shore perfectly.'
2	3	2	B	PRT	CVV.CVV	*luotoa*	*Miksi Kaisa osti postikortin?*	*Kortissa kuvattiin rannikolla olevaa luotoa täydellisesti.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'islet'	'Why did Kaisa buy the postcard?'	'The [post]card pictured the islet at the shore perfectly.'
2	3	3	N	PRT	CVVC.CVV	*luontoa*	*Kuvattiinko kortissa rannikolla olevaa kaupunkia täydellisesti?*	*Kortissa kuvattiin rannikolla olevaa luontoa täydellisesti.*
						'nature'	'Did the [post]card picture the city at the shore perfectly?'	'The [post]card pictured nature at the shore perfectly.'
2	3	1	N	PRT	CVV.CVV	*luotoa*	*Kuvattiinko kortissa rannikolla olevaa kappelia täydellisesti?*	*Kortissa kuvattiin rannikolla olevaa luotoa täydellisesti.*
						'islet'	'Did the [post]card picture the chapel at the shore perfectly?'	'The [post]card pictured the islet at the shore perfectly.'
2	3	2	B	ESS	CVVC.CV.CV	*luontona*	*Miksi Kaisa osti postikortin?*	*Kortissa oleva kuva näyttäytyi oikeana luontona täydellisesti.*
						'as nature'	'Why did Kaisa buy the postcard?'	'The picture on the [post]card perfectly looked like real nature.'
2	3	3	B	ESS	CVV.CV.CV	*luotona*	*Miksi Kaisa osti postikortin?*	*Kortissa oleva kuva näyttäytyi oikeana luotona täydellisesti.*
						'as an islet'	'Why did Kaisa buy the postcard?'	'The picture on the [post]card perfectly looked like a real islet.'
2	3	1	N	ESS	CVVC.CV.CV	*luontona*	*Näyttäytyikö kortissa oleva kuva oikeana kaupunkina täydellisesti?*	*Kortissa oleva kuva näyttäytyi oikeana luontona täydellisesti.*
						'as nature'	'Did the picture on the [post]card perfectly look like a real city?'	'The picture on the [post]card perfectly looked like real nature.'
2	3	2	N	ESS	CVV.CV.CV	*luotona*	*Näyttäytyikö kortissa oleva kuva oikeana kappelina täydellisesti?*	*Kortissa oleva kuva näyttäytyi oikeana luotona täydellisesti.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'as an islet'	'Did the picture on the [post]card perfectly look like a real chapel?'	'The picture on the [post]card perfectly looked like a real islet.'
2	4	1	B	NOM	CVVC.CV	*luonti*	*Mikä on tsättäilyssä tärkeää?*	*Ensimmäinen vaihe on foorumin luonti kotisivulla.*
						'creation/creating'	'What is important in chatting?'	'The first stage is the creation of a forum on the homepage.'
2	4	3	B	NOM	CVV.CV	*luoti*	*Miten murhatutkimus eteni?*	*Ensimmäinen todiste oli pistoolin luoti kotipihalla.*
						'bullet'	'How did the murder investigation proceed?'	'The first piece of evidence was the bullet of a pistol on the yard of the house.'
2	4	2	N	NOM	CVVC.CV	*luonti*	*Onko ensimmäinen vaihe foorumin löytäminen kotisivulla?*	*Ensimmäinen vaihe on foorumin luonti kotisivulla.*
						'creation/creating'	'The first stage is the finding the forum [lit. 'the finding of a forum'] on the homepage.'	'The first stage is the creation of a forum on the homepage.'
2	4	1	N	NOM	CVV.CV	*luoti*	*Oliko ensimmäinen todiste pistoolin hylsy kotipihalla?*	*Ensimmäinen todiste oli pistoolin luoti kotipihalla.*
						'bullet'	'Was the first piece of evidence the [projectile] shell of a pistol on the yard of the house?'	'The first piece of evidence was the bullet of a pistol on the yard of the house.'
2	4	2	B	PRT	CVV.CVV	*luotia*	*Miten murhatutkimus eteni?*	*Ensimmäinen todiste tarkoitti pistoolin luotia kotipihalla.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						‘bullet’	‘How did the murder investigation proceed?’	‘The first piece of evidence means the bullet of a pistol on the yard of the house.’
2	4	1	N	PRT	CVVC.CVV	*luontia*	*Tarkoittaako ensimmäinen vaihe foorumin löytämistä kotisivulla?*	*Ensimmäinen vaihe tarkoittaa foorumin luontia kotisivulla.*
						‘creation/creating’	‘Does the first stage mean finding a forum on the homepage?’	‘The first stage means the creation of a forum on the homepage.’
2	4	3	N	PRT	CVV.CVV	*luotia*	*Tarkoittiko ensimmäinen todiste pistoolin hylsyä kotipihalla?*	*Ensimmäinen todiste tarkoitti pistoolin luotia kotipihalla.*
						‘bullet’	‘Does the first piece of evidence mean the [projectile] shell of a pistol on the yard of the house.’	‘The first piece of evidence means the bullet of a pistol on the yard of the house.’
2	4	2	B	ESS	CVVC.CV.CV	*luontina*	*Miten tsättäily eteni?*	*Ensimmäinen vaihe toimi foorumin luontina kotisivulla.*
						‘as creation/creating’	‘How did the chat proceed?’	‘The first stage is (lit. ‘serves as’) the creation of a forum on the homepage.’
2	4	1	B	ESS	CVV.CV.CV	*luotina*	*Mitä historian opettaja kertoi?*	*Ennen lyijyn palanen toimi musketin luotina kamppailussa.*
						‘as a bullet’	‘What did the history teacher tell?’	‘Before, a small piece of lead served as the bullet of a musket in combats.’
2	4	3	N	ESS	CVVC.CV.CV	*luontina*	*Toimiko ensimmäinen vaihe foorumin löytämisenä kotisivulla?*	*Ensimmäinen vaihe toimi foorumin luontina kotisivulla.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'as creation/ creating'	'Is the first stage finding a forum (lit. 'does the first stage serves as the finding of the forum') on the homepage?'	'The first stage is (lit. 'serves as') the creation of a forum on the homepage.'
2	4	2	N	ESS	CVV.CV.CV	*luotina*	*Toimiko ennen lyijyn palanen musketin hylsynä kotipihalla?*	*Ennen lyijyn palanen toimi musketin luotina kamppailussa.*
						'as a bullet'	'Did a small piece of lead serve as the [projectile] shell of a musket in combats before?'	'Before, a small piece of lead served as the bullet of a musket in combats.'
2	5	1	B	NOM	CVV.CV	*säätö*	*Mikä oli kyselyn tulos?*	*Tunnetuin tapa säästää energiaa on lämmityksen säätö kotona.*
						'control'	'What was the result of the questionnaire?'	'The best-known way saving energy is heating control at home.'
2	5	3	N	NOM	CVVC.CV	*sääntö*	*Onko tunnetuin osa sopimusta Pekkarisen pykälä kuluista?*	*Tunnetuin osa uutta sopimusta on Pekkarisen sääntö kuluista.*
						'rule'	'Is the best-known part of the new contract Pekkarinen's clause about expenses?'	'The best-known part of the new contract is Pekkarinen's rule about expenses.'
2	5	2	N	NOM	CVV.CV	*säätö*	*Onko paras tapa säästää energia lämmityksen lopetus kotona?*	*Paras tapa säästää energiaa on lämmityksen säätö kotona.*
						'control'	'Is the best way of saving energy a heating stop at home?'	'The best way of saving energy is heating control at home.'
2	5	1	B	PRT	CVVC.CVV	*sääntöä*	*Mitä pitää ottaa huomioon?*	*Tunnetusti sopimuksessa sovelletaan Pekkarisen sääntöä kuluista.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'rule'	'What has to be taken into account?'	'It is known that the agreement applies to Pekkarinen's rule about costs.'
2	5	3	B	PRT	CVV.CVV	*säätöä*	*Mitä pitää ottaa huomioon?*	*Tunnetusti energian säästö edellyttää lämmityksen säätöä kotona.*
						'control'	'What has to be taken into account?'	'It is known that saving energy requires heating control at home.'
2	5	2	N	PRT	CVVC.CVV	*sääntöä*	*Sovelletaanko sopimuksessa tunnetusti Pekkarisen pykälää kuluista?*	*Tunnetusti sopimuksessa sovelletaan Pekkarisen sääntöä kuluista.*
						'rule'	'Is it known that the agreement applies to Pekkarinen's clause about costs?'	'It is known that the agreement applies to Pekkarinen's rule about costs.'
2	5	1	N	PRT	CVV.CVV	*säätöä*	*Edellyttääkö energian säästö tunnetusti lämmityksen lopetusta kotona?*	*Tunnetusti energian säästö edellyttää lämmityksen säätöä kotona.*
						'control'	'Is it known that saving energy requires a heating stop at home?'	'It is known that saving energy requires heating control control at home.'
2	5	3	B	ESS	CVVC.CV.CV	*sääntönä*	*Mikä oli kyselyn tulos?*	*Parhaiten uusi sopimus tunnetaan Pekkarisen sääntönä kuluista.*
						'as (a) control'	'What was the result of the questionnaire?'	'The new agreement is best-known as Pekkarinen's rule about costs.'
2	5	2	B	ESS	CVV.CV.CV	*säätönä*	*Mikä oli kyselyn tulos?*	*Parhaiten energian säästö ilmenee lämmityksen säätönä kotona.*
						'as a rule'	'What was the result of the questionnaire?'	'Saving energy is best instantiated as heating control at home.'
2	5	1	N	ESS	CVVC.CV.CV	*sääntönä*	*Tunnetaanko uusi sopimus parhaiten Pekkarisen pykälänä kuluista?*	*Parhaiten uusi sopimus tunnetaan Pekkarisen sääntönä kuluista.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'as (a) control'	'Is the new agreement best-known as Pekkarinen's clause about costs.'	'The new agreement is best-known as Pekkarinen's rule about costs.'
2	5	3	N	ESS	CVV.CV.CV	*säätönä*	*Ilmeneekö energian säästö parhaiten lämmityksen lopetuksena kotona?*	*Parhaiten energian säästö ilmenee lämmityksen säätönä kotona.*
						'as a rule'	'Is saving energy best instantiated as a heating stop at home?'	'Saving energy is best instantiated as heating control at home.'
2	6	3	B	NOM	CVVC.CV	*kiilto*	*Miksi Veera luki biologiaa?*	*Veeran hurmasi aina sudenkorentojen kiilto kesäyössä.*
						'glint(ing)'	'Why did Veera study biology?'	'Veera was always fascinated by dragonflies' glinting on a summer night.'
2	6	2	B	NOM	CVV.CV	*kiito*	*Miksi Veera luki biologiaa?*	*Veeran hurmasi aina sudenkorentojen kiito kesäyössä.*
						'zoom(ing)'	'Why did Veera study biology?'	'Veera was always fascinated by dragonflies' zooming on a summer night.'
2	6	1	N	NOM	CVVC.CV	*kiilto*	*Hurmasiko Veeran aina sudenkorentojen suuruus kesäyössä?*	*Veeran hurmasi aina sudenkorentojen kiilto kesäyössä.*
						'glint(ing)'	'Was Veera always fascinated by dragonflies' size on a summer night?'	'Veera was always fascinated by dragonflies' glinting on a summer night.'
2	6	3	N	NOM	CVV.CV	*kiito*	*Hurmasiko Veeran aina sudenkorentojen suuruus kesäyössä?*	*Veeran hurmasi aina sudenkorentojen kiito kesäyössä.*
						'zoom(ing)'	'Was Veera always fascinated by dragonflies' size on a summer night?'	'Veera was always fascinated by dragonflies' zooming on a summer night.'
2	6	2	B	PRT	CVVC.CVV	*kiiltoa*	*Mistä puhuttiin?*	*Vera kehui hienosti sudenkorentojen kiiltoa kesäyössä.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'glint(ing)'	'What did you talk about?'	'Veera nicely praised dragonflies' glinting on a summer night.'
2	6	1	B	PRT	CVV.CVV	*kiitoa*	*Mistä puhuttiin?*	*Vera kehui hienosti sudenkorentojen kiitoa kesäyössä.*
						'zoom(ing)'	'What did you talk about?'	'Veera nicely praised dragonflies' zooming on a summer night.'
2	6	3	N	PRT	CVVC.CVV	*kiiltoa*	*Kehuiko Veera hienosti sudenkorentojen suuruutta kesäyössä?*	*Vera kehui hienosti sudenkorentojen kiiltoa kesäyössä.*
						'glint(ing)'	'Did Veera nicely praise dragonflies' size on a summer night.'	'Veera nicely praised dragonflies' glinting on a summer night.'
2	6	2	N	PRT	CVV.CVV	*kiitoa*	*Kehuiko Veera hienosti sudenkorentojen suuruutta kesäyössä?*	*Vera kehui hienosti sudenkorentojen kiitoa kesäyössä.*
						'zoom(ing)'	'Did Veera nicely praise dragonflies' size on a summer night.'	'Veera nicely praised dragonflies' zooming on a summer night.'
2	6	1	B	ESS	CVVC.CV.CV	*kiiltona*	*Mistä oli puhetta?*	*Veera kuvaili näkemäänsä sudenkorentojen kiiltona kesäyössä.*
						'as a glint'	'What was the topic?'	'Veera described how she saw dragonflies as a glint on a summer night.'
2	6	3	B	ESS	CVV.CV.CV	*kiitona*	*Mistä oli puhetta?*	*Veera kuvaili näkemäänsä sudenkorentojen kiitona kesäyössä.*
						'as a zoom'	'What was the topic?'	'Veera described how she saw dragonflies as a zoom on a summer night.'
2	6	2	N	ESS	CVVC.CV.CV	*kiiltona*	*Kuvailiko Veera näkemäänsä sudenkorentojen suhinana kesäyössä?*	*Veera kuvaili näkemäänsä sudenkorentojen kiiltona kesäyössä.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'as a glint'	'Did Veera describe how she saw dragonflies as a whoosh on a summer night?'	'Veera described how she saw dragonflies as a glint on a summer night.'
2	6	1	N	ESS	CVV.CV.CV	*kiitona*	*Kuvailiko Veera näkemäänsä sudenkorentojen lentelynä kesäyössä?*	*Veera kuvaili näkemäänsä sudenkorentojen kiitona kesäyössä.*
						'as a zoom'	'Did Veera describe how she saw dragonflies flying around on a summer night?'	'Veera described how she saw dragonflies as a zoom on a summer night.'
3	1	1	B	NOM	CVVC.CV	*paarma*	*Miksi Marjut on yllättynyt?*	*Hannun ottamassa kuvassa näkyy ensisilmäyksellä paarma mustavalkoisena.*
						'horsefly'	'Why is Marjut surprised?'	'At first glance, the picture that Hannu took shows a horsefly in black and white.'
3	1	2	B	NOM	CVC.CV	*Parma*	*Miksi Marjut on yllättynyt?*	*Hannun ottamassa kuvassa näkyy ensisilmäyksellä Parma mustavalkoisena.*
						'Parma (city in Italy)'	'Why is Marjut surprised?'	'At first glance, the picture that Hannu took shows Parma in black and white.'
3	1	3	N	NOM	CVVC.CV	*paarma*	*Näkyykö Hannun ottamassa kuvassa ensisilmäyksellä kärpänen mustavalkoisena?*	*Hannun ottamassa kuvassa näkyy ensisilmäyksellä paarma mustavalkoisena.*
						'horsefly'	'Does the picture that Hannu took at first glance show a fly in black and white?'	'At first glance, the picture that Hannu took shows a horsefly in black and white.'
3	1	1	N	NOM	CVC.CV	*Parma*	*Näkyykö Hannun ottamassa kuvassa ensisilmäyksellä Rooma mustavalkoisena?*	*Hannun ottamassa kuvassa näkyy ensisilmäyksellä Parma mustavalkoisena.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						‘Parma (city in Italy)’	‘Does the picture that Hannu took at first glance show Rome in black and white?’	‘At first glance, the picture that Hannu took shows Parma in black and white.’
3	1	2	B	PRT	CVVC.CVV	*paarmaa*	*Miksi Marjut on yllättynyt?*	*Hannun ottamassa videossa kuvataan ensisilmäyksellä paarmaa mustavalkoisena.*
						‘horsefly’	‘Why is Marjut surprised?’	‘At first glance, the video that Hannu took depicts a horselfy in black and white.’
3	1	3	B	PRT	CVC.CVV	*Parmaa*	*Miksi Marjut on yllättynyt?*	*Hannun ottamassa videossa kuvataan ensisilmäyksellä Parmaa mustavalkoisena.*
						‘Parma (city in Italy)’	‘Why is Marjut surprised?’	‘At first glance, the video that Hannu took depicts Parma in black and white.’
3	1	1	N	PRT	CVVC.CVV	*paarmaa*	*Kuvataanko Hannun ottamassa videossa ensisilmäyksellä kärpänen mustavalkoisena?*	*Hannun ottamassa videossa kuvataan ensisilmäyksellä paarmaa mustavalkoisena.*
						‘horsefly’	‘Does the video that Hannu took at first glance depict a fly in black and white?’	‘At first glance, the video that Hannu took depicts a horselfy in black and white.’
3	1	2	N	PRT	CVC.CVV	*Parmaa*	*Kuvataanko Hannun ottamassa videossa ensisilmäyksellä Rooma mustavalkoisena?*	*Hannun ottamassa videossa kuvataan ensisilmäyksellä Parmaa mustavalkoisena.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'Parma (city in Italy)'	'Does the video that Hannu took at first glance depict Rome in black and white?'	'At first glance, the video that Hannu took depicts Parma in black and white.'
3	1	3	B	ESS	CVVC.CV.CV	*paarmana*	*Miksi Marjut on yllättynyt?*	*Hannu piti kuvassa olevaa ötökkää ensisilmäyksellä paarmana mustikanlehdellä.*
						'as a horsefly'	'Why is Marjut surprised?'	'At first glance, Hannu thought the bug on the picture was a horsefly in black and white.'
3	1	1	B	ESS	CVC.CV.CV	*Parmana*	*Miksi Marjut on yllättynyt?*	*Hannu piti kuvassa olevaa kaupunkia ensisilmäyksellä Parmana musiikkiviikoilla.*
						'as Parma (city in Italy)'	'Why is Marjut surprised?'	'At first glance, Hannu thought the city on the picture was Parma in black and white.'
3	1	2	N	ESS	CVVC.CV.CV	*paarmana*	*Pitikö Hannu kuvassa olevaa ötökkää ensisilmäyksellä kärpäsenä mustikanlehdellä?*	*Hannu piti kuvassa olevaa ötökkää ensisilmäyksellä paarmana mustikanlehdellä.*
						'as a horsefly'	'Did Hannu at first glance think the bug on the picture was a fly in black and white?'	'At first glance, Hannu thought the bug on the picture was a horsefly in black and white.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
3	1	3	N	ESS	CVC.CV.CV	*Parmana*	*Pitikö Hannu kuvassa olevaa kaupunkia ensisilmäyksellä Roomana musiikkivi-ikoilla?*	*Hannu piti kuvassa olevaa kaupunkia ensisilmäyksellä Parmana musiikkivi-ikoilla.*
						'as Parma (city in Italy)'	'Did Hannu at first glance think the city on the picture was Rome in black and white?'	'At first glance, Hannu thought the city on the picture was Parma in black and white.'
3	2	2	B	NOM	CVVC.CV	*vaarna*	*Mitä opit lomalla?*	*Teltan tärkein osa on vaarna tietääkseni.*
						'tent peg'	'What did you learn during your holi-days?'	'The most important part of a tent is the tent peg, as far as I know.'
3	2	3	B	NOM	CVC.CV	*Varna*	*Mitä opit lomalla?*	*Bulgarian kaunein kaupunki on Varna tietääkseni.*
						'Varna (city in Bul-garia)'	'What did you learn during your holi-days?'	'The most beautiful city in Bulgaria is Varna, as far as I know.'
3	2	1	N	NOM	CVVC.CV	*vaarna*	*Onko teltan tärkein osa naru tietääk-sesi?*	*Teltan tärkein osa on vaarna tietääkseni.*
						'tent peg'	'Is the most important part of a tent the cord, as far as you know?'	'The most important part of a tent is the tent peg, as far as I know.'
3	2	2	N	NOM	CVC.CV	*Varna*	*Onko Bulgarian kauniin kaupunki Sofia tietääksesi?*	*Bulgarian kaunein kaupunki on Varna tietääkseni.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'Varna (city in Bulgaria)'	'Is the most beautiful city in Bulgaria is Sofia, as far as you know?'	'The most beautiful city in Bulgaria is Varna, as far as I know.'
3	2	3	B	PRT	CVVC.CVV	*vaarnaa*	*Kuinka säästettiin rahaa?*	*Kesällä tehtiin lankusta vaarnaa tietääkseni.*
						'tent peg'	'How did they save money?'	'In the summer, they made a tent peg from a plank, as far as I know.'
3	2	1	B	PRT	CVC.CVV	*Varnaa*	*Kuinka säästettiin rahaa?*	*Leffassa tehtiin Riikasta Varnaa tietääkseni.*
						'Varna (city in Bulgaria)'	'How did they save money?'	'In the movie, they made Riga into Varna, as far as I know.'
3	2	2	N	PRT	CVVC.CVV	*vaarnaa*	*Tehtiinkö kesällä lankusta hyllyä tietääksesi?*	*Kesällä tehtiin lankusta vaarnaa tietääkseni.*
						'tent peg'	'Did they made shelves from a plank in the summer as far as I know?'	'In the summer, they made a tent peg from a plank, as far as I know.'
3	2	3	N	PRT	CVC.CVV	*Varnaa*	*Tehtiinkö leffassa Riikasta Helsinkiä tietääksesi?*	*Leffassa tehtiin Riikasta Varnaa tietääkseni.*
						'Varna (city in Bulgaria)'	'Did they made Riga into Varna in the movie, as far as you know?'	'In the movie, they made Riga into Varna, as far as I know.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
3	2	1	B	ESS	CVVC.CV.CV	*vaarnana*	*Kuinka säästettiin rahaa?*	*Kesällä käytettiin rimaa vaarnana tietääkseni.*
						'as a tent peg'	'How did they save money?'	'In the summer, they used a lath as a tent peg, as far as I know.'
3	2	2	B	ESS	CVC.CV.CV	*Varnana*	*Kuinka säästettiin rahaa?*	*Leffassa filmattiin Riikaa Varnana tietääkseni.*
						'as Varna (city in Bulgaria)'	'How did they save money?'	'In the movie, they filmed Riga instead of [lit. 'as'] Varna, as far as I know.'
3	2	3	N	ESS	CVVC.CV.CV	*vaarnana*	*Käytettiinkö kesällä rimaa onkena tietääksesi?*	*Kesällä käytettiin rimaa vaarnana tietääkseni.*
						'as a tent peg'	'Did they use a lath as a fishing rod in the summer, as far as you know?'	'In the summer, they used a lath as a tent peg, as far as I know.'
3	2	1	N	ESS	CVC.CV.CV	*Varnana*	*Filmattiinko leffassa Riikaa Helsinkinä tietääksesi?*	*Leffassa filmattiin Riikaa Varnana tietääkseni.*
						'as Varna (city in Bulgaria)'	'Did they film Riga instead of [lit. 'as'] Helsinki in the movie, as far as you know?'	'In the movie, they filmed Riga instead of [lit. 'as'] Varna, as far as I know.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
3	3	3	B	NOM	CVVC.CV	*Haarma*	*Mitä oli tietokilpailussa oikea vastaus?*	*Tunnetun musiikkitoimittajan nimi on tietenkin Haarma tällä kertaa.*
						'Haarma (last name)'	'What was the correct answer in the quiz?'	'The well-known music journalist's name is, of course, Haarma this time.'
3	3	1	B	NOM	CVC.CV	*Harma*	*Mitä oli tietokilpailussa oikea vastaus?*	*Tunnetun musiikkitoimittajan nimi on tietenkin Harma tällä kertaa.*
						'Harma (last name)'	'What was the correct answer in the quiz?'	'The well-known music journalist's name is, of course, Harma this time.'
3	3	2	N	NOM	CVVC.CV	*Haarma*	*Onko tunnetun musiikkitoimittajan nimi Holopainen tällä kertaa?*	*Tunnetun musiikkitoimittajan nimi on tietenkin Haarma tällä kertaa.*
						'Haarma (last name)'	'Was the well-known music journalist's name Holopainen this time?'	'The well-known music journalist's name is, of course, Haarma this time.'
3	3	3	N	NOM	CVC.CV	*Harma*	*Onko tunnetun musiikkitoimittajan nimi Holopainen tällä kertaa?*	*Tunnetun musiikkitoimittajan nimi on tietenkin Harma tällä kertaa.*
						'Harma (last name)'	'Was the well-known music journalist's name Holopainen this time?'	'The well-known music journalist's name is, of course, Harma this time.'
3	3	1	B	PRT	CVVC.CVV	*Haarmaa*	*Miten best-of-listan tekeminen meni?*	*Tunnetuimmaksi musiikkitoimittajaksi ehdotettiin tietenkin Haarmaa tällä kertaa.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'Haarma (last name)'	'How did the making of the best-of-list go?'	'As the best-known music journalist, of course, Haarma was suggested this time.'
3	3	2	B	PRT	CVC.CVV	*Harmaa*	*Miten best-of-listan tekeminen meni?*	*Tunnetuimmaksi musiikkitoimittajaksi ehdotettiin tietenkin Harmaa tällä kertaa.*
						'Harma (last name)'	'How did the making of the best-of-list go?'	'As the best-known music journalist, of course, Harma was suggested this time.'
3	3	3	N	PRT	CVVC.CVV	*Haarmaa*	*Ehdotettiinko tunnetuimmaksi musiikkitoimittajaksi Holopaista tällä kertaa?*	*Tunnetuimmaksi musiikkitoimittajaksi ehdotettiin tietenkin Haarmaa tällä kertaa.*
						'Haarma (last name)'	'Was Holopainen suggested as the best-known music journalist this time?'	'As the best-known music journalist, of course, Haarma was suggested this time.'
3	3	1	N	PRT	CVC.CVV	*Harmaa*	*Ehdotettiinko tunnetuimmaksi musiikkitoimittajaksi Holopaista tällä kertaa?*	*Tunnetuimmaksi musiikkitoimittajaksi ehdotettiin tietenkin Harmaa tällä kertaa.*
						'Harma (last name)'	'Was Holopainen suggested as the best-known music journalist this time?'	'As the best-known music journalist, of course, Harma was suggested this time.'
3	3	2	B	ESS	CVVC.CV.CV	*Haarmana*	*Miten best-of-listan tekeminen meni?*	*Tunnettua musiikkitoimittajaa pidettiin tietenkin Haarmana tällä kertaa.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'as Haarma (last name)'	'How did the making of the best-of-list go?'	'Of course, Haarma was considered the well-known music journalist this time.'
3	3	3	B	ESS	CVC.CV.CV	*Harmana*	*Miten best-of-listan tekeminen meni?*	*Tunnettua musiikkitoimittajaa pidettiin tietenkin Harmana tällä kertaa.*
						'as Harma (last name)'	'How did the making of the best-of-list go?'	'Of course, Harma was considered the well-known music journalist this time.'
3	3	1	N	ESS	CVVC.CV.CV	*Haarmana*	*Pidettiinkö tunnettua musiikkitoimittajaa Holopaisena täällä kertaa?*	*Tunnettua musiikkitoimittajaa pidettiin tietenkin Haarmana tällä kertaa.*
						'as Haarma (last name)'	'Was Holopainen considered the well-known music journalist this time?'	'Of course, Haarma was considered the well-known music journalist this time.'
3	3	2	N	ESS	CVC.CV.CV	*Harmana*	*Pidettiinkö tunnettua musiikkitoimittajaa Holopaisena täällä kertaa?*	*Tunnettua musiikkitoimittajaa pidettiin tietenkin Harmana tällä kertaa.*
						'as Harma (last name)'	'Was Holopainen considered the well-known music journalist this time?'	'Of course, Harma was considered the well-known music journalist this time.'
3	4	1	B	NOM	CVVC.CV	*tuonti*	*Kuinka hän tiesi asiasta?*	*Pekalle oli selvää missä ilmoitetaan alkoholin tuonti tänä syksynä.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'import'	'How did (s)he know about the issue?'	'To Pekka, it was clear where to report the alcohol import this autumn.'
3	4	3	B	NOM	CVC.CV	*tunti*	*Kuinka hän tiesi asiasta?*	*Pekalle oli selvää missä pidetään englannin tunti tänä syksynä.*
						'hour'	'How did (s)he know about the issue?'	'To Pekka, it was clear where to conduct the English lesson this autumn.'
3	4	2	N	NOM	CVVC.CV	*tuonti*	*Oliko Pekalle selvää missä ilmoitetaan alkoholin myynti tänä syksynä?*	*Pekalle oli selvää missä ilmoitetaan alkoholin tuonti tänä syksynä.*
						'import'	'Was it clear to Pekka where to report the alcohol sale this autumn?'	'To Pekka, it was clear where to report the alcohol import this autumn.'
3	4	1	N	NOM	CVC.CV	*tunti*	*Oliko Pekalle selvää missä ilmoitetaan pidetään englannin koe tänä syksynä?*[2]	*Pekalle oli selvää missä pidetään englannin tunti tänä syksynä.*
						'hour'	'Was it clear to Pekka where to report the English test this autumn?'	'To Pekka, it was clear where to conduct the English lesson this autumn.'
3	4	3	B	PRT	CVVC.CVV	*tuontia*	*Kuinka hän tiesi asiasta?*	*Pekalle oli selvää välttää alkoholin tuontia tänä syksynä.*
						'import'	'How did (s)he know about the issue?'	'To Pekka, it was clear to avoid alcohol import this autumn.'
3	4	2	B	PRT	CVC.CVV	*tuntia*	*Kuinka hän tiesi asiasta?*	*Pekalle oli selvää välttää englannin tuntia tänä syksynä.*

Continued on next page

[2]This prompt contained a superfluous word *ilmoitetaan* 'one reports' and thus two (passive) verbs. However, this did not obstruct the intended focus structure, so that fluently uttered responses were all retained for analysis.

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'hour'	'How did (s)he know about the issue?'	'To Pekka, it was clear to avoid the English lesson this autumn.'
3	4	1	N	PRT	CVVC.CVV	*tuontia*	*Oliko Pekalle selvää välttää alkoholin juontia tänä syksynä?*	*Pekalle oli selvää välttää alkoholin tuontia tänä syksynä.*
						'import'	'Was it clear to Pekka to avoid alcohol consumption [lit. 'drinking of alcohol'] this autumn.'	'To Pekka, it was clear to avoid alcohol import this autumn.'
3	4	3	N	PRT	CVC.CVV	*tuntia*	*Oliko Pekalle selvää välttää englannin koetta tänä syksynä?*	*Pekalle oli selvää välttää englannin tuntia tänä syksynä.*
						'hour'	'Was it clear to Pekka to avoid the English test this autumn?'	'To Pekka, it was clear to avoid the English lesson this autumn.'
3	4	2	B	ESS	CVVC.CV.CV	*tuontina*	*Kuinka asia tuli esiin?*	*Tullille oli itsestään selvää pitää sitä alkoholin tuontina tänä syksynä.*
						'as import'	'How the issue become apparent?'	'It was obvious to customs to view this as alcohol import this autumn.'
3	4	1	B	ESS	CVC.CV.CV	*tuntina*	*Kuinka asia tuli esiin?*	*Opettajalle oli itsestään selvää pitää se englannin tuntina tänä syksynä.*
						'as an hour'	'How the issue become apparent?'	'It was obvious to the teacher to view this as an English lesson this autumn.'
3	4	3	N	ESS	CVVC.CV.CV	*tuontina*	*Oliko tullille selvää pitää asiaa alkoholin myyntinä tänä syksynä?*	*Tullille oli itsestään selvää pitää sitä alkoholin tuontina tänä syksynä.*
						'as import'	'Was it obvious to customs to view this as alcohol sales this autumn?'	'It was obvious to customs to view this as alcohol import this autumn.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
3	4	2	N	ESS	CVC.CV.CV	*tuntina*	*Oliko opettajalle selvää pitää opetus englannin kokeena tänä syksynä?*	*Opettajalle oli itsestään selvää pitää se englannin tuntina tänä syksynä.*
						'as an hour'	'Was it obvious to the teacher to give a lecture as an English test this autumn?'	'It was obvious to the teacher to teach this as an English lesson this autumn.'
3	5	2	B	NOM	CVVC.CV	*hionta*	*Kuinka tiesit, että pajassa oli joku?*	*Sisään tullessa kuului edeltä hionta selvästi.*
						'grinding/ polish-ing'	'How did you know that sombody was in the workshop?'	'When coming inside, [the sound of] grinding was clearly audible before.'
3	5	1	B	NOM	CVC.CV	*hinta*	*Kuinka tiesit, että kaupassa ei oltu käyty?*	*Vanhoissa tuotteissa oli vielä hinta selvästi.*
						'price'	'How did you know that nobody had been in the shop?'	'On the old products, there was clearly still the price.'
3	5	3	N	NOM	CVVC.CV	*hionta*	*Kuuluiko pajaan tullessa edeltä sahaus selvästi?*	*Sisään tullessa kuului edeltä hionta selvästi.*
						'grinding/ polish-ing'	'When coming inside, was [the sound of] sawing was clearly audible before?'	'When coming inside, [the sound of] grinding was clearly audible before.'
3	5	2	N	NOM	CVC.CV	*hinta*	*Oliko vanhoissa tuotteissa vielä viivakoodi selvästi?*	*Vanhoissa tuotteissa oli vielä hinta selvästi.*
						'price'	'Was there was clearly still a bar code on the old products?'	'On the old products, there was clearly still a price.'
3	5	1	B	PRT	CVVC.CVV	*hiontaa*	*Kuinka tiesit, että pajassa oli joku?*	*Sisään tullessa kuului edeltä hiontaa selvästi.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'grinding/ polishing'	'How did you know that sombody was in the workshop?'	'When coming inside, [the sound of] grinding was clearly audible before.'
3	5	3	B	PRT	CVC.CVV	*hintaa*	*Kuinka tiesit, että kaupassa oli käyty?*	*Vanhoissa tuotteissa oli muutettu vielä hintaa selvästi.*
						'price'	'How did you know that somebody had been in the shop?'	'On the old products, the price was still clearly changed.'
3	5	2	N	PRT	CVVC.CVV	*hiontaa*	*Kuuluiko pajaan tullessa edeltä sahaus selvästi?*	*Sisään tullessa kuului edeltä hiontaa selvästi.*
						'grinding/ polishing'	'When coming inside, was [the sound of] sawing was clearly audible before?'	'When coming inside, [the sound of] grinding was clearly audible before.'
3	5	1	N	PRT	CVC.CVV	*hintaa*	*Oliko vanhoissa tuotteissa muutettu vielä viivakoodia selvästi?*	*Vanhoissa tuotteissa oli muutettu vielä hintaa selvästi.*
						'price'	'Was the bar code clearly still changed on the old products?'	'On the old products, the price was still clearly changed.'
3	5	3	B	ESS	CVVC.CV.CV	*hiontana*	*Miksi oppipoika teki sen sillä tavalla?*	*Tämä oli työprosessissa toiminnut edeltä hiontana selvästi.*
						'as grinding/ polishing'	'Why did the apprentice do it that way?'	'This had clearly functioned as the grinding in the work process before.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
3	5	2	B	ESS	CVC.CV.CV	*hintana*	*Miksi otit niin paljon rahaa mukaan?*	*Tämä summa oli toiminnut edeltä hintana selvästi.*
						'as a price'	'Why did you take that much money with you?'	'This sum had clearly functioned as the price before.'
3	5	1	N	ESS	CVVC.CV.CV	*hiontana*	*Oliko tämä toiminnut työprosessissa edeltä sahauksena selvästi?*	*Tämä oli työprosessissa toiminnut edeltä hiontana selvästi.*
						'as grinding/ polishing'	'Had this clearly functioned as the sawing in the work process before?'	'This had clearly functioned as the grinding in the work process before.'
3	5	3	N	ESS	CVC.CV.CV	*hintana*	*Oliko tämä summa toiminnut edeltä ehdotuksena selvästi?*	*Tämä summa oli toiminnut edeltä hintana selvästi.*
						'as a price'	'Had this sum had clearly functioned a suggestion before?'	'This sum had clearly functioned as the price before.'
3	6	3	B	NOM	CVVC.CV	*syönti*	*Mistä luostarilla puhuttiin?*	*Munkkien lempiaihe oli syönti paastoaikana.*
						'eating'	'What were they talking about in the monastry?'	'The monks' favourite topic was eating during the fasting time.'
3	6	2	B	NOM	CVC.CV	*synti*	*Mistä luostarilla puhuttiin?*	*Munkkien lempiaihe oli synti paastoaikana.*
						'sin'	'What were they talking about in the monastry?'	'The monks' favourite topic was the sin during the fasting time.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
3	6	1	N	NOM	CVVC.CV	*syönti*	*Oliko munkkien lempiaihe juonti paastoaikana?*	*Munkkien lempiaihe oli syönti paastoaikana.*
						'eating'	'Was the monks' favourite topic drinking during the fasting time?'	'The monks' favourite topic was eating during the fasting time.'
3	6	3	N	NOM	CVC.CV	*synti*	*Oliko munkkien lempiaihe armahdus paastoaikana?*	*Munkkien lempiaihe oli synti paastoaikana.*
						'sin'	'Was the monks' favourite topic amnesty during the fasting time?'	'The monks' favourite topic was the sin during the fasting time.'
3	6	2	B	PRT	CVVC.CVV	*syöntiä*	*Mitä paikalliset kertovat luolasta?*[3]	*Munkkien lempiaihe käsitteli syöntiä paastoaikana.*
						'eating'	'What did the locals tell about the cave?'	'The monks' favourite topic dealt with eating during the fasting time.'
3	6	1	B	PRT	CVC.CVV	*syntiä*	*Mitä paikalliset kertovat luolasta?*[4]	*Munkkien lempiaihe käsitteli syntiä paastoaikana.*
						'sin'	'What did the locals tell about the cave?'	'The monks' favourite topic dealt with the sin during the fasting time.'
3	6	3	N	PRT	CVVC.CVV	*syöntiä*	*Käsittelikö munkkien lempiaihe juontia paastoaikana?*	*Munkkien lempiaihe käsitteli syöntiä paastoaikana.*
						'eating'	'Did the monks' favourite topic deal with drinking during the fasting time?'	'The monks' favourite topic dealt with eating during the fasting time.'

Continued on next page

[3]This question should have contained the word *luostarista* 'about the monastery' instead of *luolasta* 'about the cave' to be congruent with the answer. However, since the intended focus structure was broad (all-new), resulting materials were retained.

[4]This prompt question contains the same mistake as the previous one and answers were retained following the same rationale.

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
3	6	2	N	PRT	CVC.CVV	*syntiä*	*Käsittelikö munkkien lempiaihe armahdusta paastoaikana?*	*Munkkien lempiaihe käsitteli syntiä paastoaikana.*
						'sin'	'Did the monks' favourite topic deal with amnesty during the fasting time?'	'The monks' favourite topic dealt with the sin during the fasting time.'
3	6	1	B	ESS	CVVC.CV.CV	*syöntinä*	*Kuulitko jotakin yllättävää?*	*Keskiajalla munkeille olut kävi syöntinä paastoaikana.*
						'as eating/food'	'Did you hear something surprising?'	'In the middle ages, beer served as food for the munks during the fasting time.'
3	6	3	B	ESS	CVC.CV.CV	*syntinä*	*Kuulitko jotakin yllättävää?*	*Keskiajalla munkeille olut kävi syntinä paastoaikana.*
						'as a sin'	'Did you hear something surprising?'	'In the middle ages, beer went as a sin for the munks during the fasting time.'
3	6	2	N	ESS	CVVC.CV.CV	*syöntinä*	*Kävikö keskiajalla munkeille olut kauppatavarana paastoaikana?*	*Keskiajalla munkeille olut kävi syöntinä paastoaikana.*
						'as eating/food'	'In the middle ages, did beer serve as a commodity for the munks during the fasting time?'	'In the middle ages, beer served as food for the munks during the fasting time.'
3	6	1	N	ESS	CVC.CV.CV	*syntinä*	*Kävikö keskiajalla munkeille olut viihteenä paastoaikana?*	*Keskiajalla munkeille olut kävi syntinä paastoaikana.*
						'as a sin'	'In the middle ages, did beer go as entertainment for the munks during the fasting time?'	'In the middle ages, beer went as a sin for the munks during the fasting time.'
4	1	1	B	NOM	CVC.CV	*kulma*	*Miksi riitelitte?*	*Pöydälle paras paikka oli kulma tällä kertaa.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'corner'	'Why did you fight?'	'The best place for the table was the corner this time.'
4	1	2	B	NOM	CVCg.CV	*kumma*	*Miksi riitelitte?*	*Katrista pöydän paikka oli kumma tällä kertaa.*
						'odd'	'Why did you fight?'	'To Katri the place of the table was odd this time.'
4	1	3	N	NOM	CVC.CV	*kulma*	*Onko pöydälle paras paikka ovensuu tällä kertaa?*	*Pöydälle paras paikka oli kulma tällä kertaa.*
						'corner'	'Was the best place for the table the doorway this time?'	'The best place for the table was the corner this time.'
4	1	1	N	NOM	CVCg.CV	*kumma*	*Onko Katrista pöydän paikka hyvä tällä kertaa?*	*Katrista pöydän paikka oli kumma tällä kertaa.*
						'odd'	'Was the place of the table was good [according] to Katri this time?'	'To Katri the place of the table was odd this time.'
4	1	2	B	PRT	CVC.CVV	*kulmaa*	*Miksi riitelitte?*	*Katrista pöydän muoto koristeli kulmaa tällä kertaa.*
						'corner'	'Why did you fight?'	'To Katri the shape of the table decorated the corner this time.'
4	1	3	B	PRT	CVCg.CVV	*kummaa*	*Miksi riitelitte?*	*Katrista pöydän heilunta oli kummaa tällä kertaa.*
						'odd'	'Why did you fight?'	'To Katri the sway of the table was odd this time.'
4	1	1	N	PRT	CVC.CVV	*kulmaa*	*Koristeliko Katrista pöydän muoto huonetta tällä kertaa?*	*Katrista pöydän muoto koristeli kulmaa tällä kertaa.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'corner'	'Did the shape of the table decorate the room to Katri this time?'	'To Katri the shape of the table decorated the corner this time.'
4	1	2	N	PRT	CVCg.CVV	*kummaa*	*Oliko Katrista pöydän heilunta kivaa tällä kertaa?*	*Katrista pöydän heilunta oli kummaa tällä kertaa.*
						'odd'	'Was the sway of the table was nice [according] to Katri this time?'	'To Katri the sway of the table was odd this time.'
4	1	3	B	ESS	CVC.CV.CV	*kulmana*	*Miksi riitelitte?*	*Katrista umpikuja ilmeni kulmana tällä kertaa.*
						'as a corner'	'Why did you fight?'	'To Katri the dead end appeared as a corner this time.'
4	1	1	B	ESS	CVCg.CV.CV	*kummana*	*Miksi riitelitte?*	*Katria pöydän malli häiritsi kummana tällä kertaa.*
						'as odd'	'Why did you fight?'	'Katri was bothered by the table's model being odd this time.'
4	1	2	N	ESS	CVC.CV.CV	*kulmana*	*Ilmenikö Katrista umpikuja porttina tällä kertaa?*	*Katrista umpikuja ilmeni kulmana tällä kertaa.*
						'as a corner'	'Did the dead end appear as a gate to Katri this time?'	'To Katri the dead end appeared as a corner this time.'
4	1	3	N	ESS	CVCg.CV.CV	*kummana*	*Häiritsikö Katria pöydän malli halpana tällä kertaa?*	*Katria pöydän malli häiritsi kummana tällä kertaa.*
						'as odd'	'Was Katri bothered by the table's model being cheap this time?'	'Katri was bothered by the table's model being odd this time.'
4	2	2	B	NOM	CVC.CV	*silmu*	*Mitä sinun opettaja-kaveri kertoi tällä viikolla?*	*Taideluokassa aiheena oli mieluiten silmu piirustuksessa.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'bud of a leaf'	'What did your teacher friend tell this week?'	'In art class, the subject was mostly the bud of a leaf in drawing.'
4	2	3	B	NOM	CVCg.CV	*simmu*	*Mitä sinun opettaja-kaveri kertoi tällä viikolla?*	*Taideluokassa aiheena oli mieluiten simmu piirustuksessa.*
						'eye'	'What did your teacher friend tell this week?'	'In art class, the subject was mostly the eye in drawing.'
4	2	1	N	NOM	CVC.CV	*silmu*	*Oliko taideluokassa aiheena mieluiten taivas piirustuksessa?*	*Taideluokassa aiheena oli mieluiten silmu piirustuksessa.*
						'bud of a leaf'	'Was the subject in art class mostly the sky in drawing?'	'In art class, the subject was mostly the bud of a leaf in drawing.'
4	2	2	N	NOM	CVCg.CV	*simmu*	*Oliko taideluokassa aiheena mieluiten taivas piirustuksessa?*	*Taideluokassa aiheena oli mieluiten simmu piirustuksessa.*
						'eye'	'Was the subject in art class mostly the sky in drawing?'	'In art class, the subject was mostly the eye in drawing.'
4	2	3	B	PRT	CVC.CVV	*silmua*	*Miksi luulet, että isän lempitaitelija oli outo kaveri?*	*Taitelija kuvasi nuorena mieluiten silmua piirustuksessa.*
						'bud of a leaf'	'Why do you think that father's favourite artist is a strange guy?'	'When the artist was young, (s)he mostly depicted the bud of a leaf in drawing.'
4	2	1	B	PRT	CVCg.CVV	*simmua*	*Miksi luulet, että isän lempitaitelija oli outo kaveri?*	*Taitelija kuvasi nuorena mieluiten simmua piirustuksessa.*
						'eye'	'Why do you think that father's favourite artist is a strange guy?'	'When the artist was young, (s)he mostly depicted the eye in drawing.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
4	2	2	N	PRT	CVC.CVV	*silmua*	*Kuvasiko taitelija nuorena mieluiten taivasta piirustuksessa?*	*Taitelija kuvasi nuorena mieluiten silmua piirustuksessa.*
						'bud of a leaf'	'When the artist was young, did (s)he mostly depicted the sky in drawing?'	'When the artist was young, (s)he mostly depicted the bud of a leaf in drawing.'
4	2	3	N	PRT	CVCg.CVV	*simmua*	*Kuvasiko taitelija nuorena mieluiten taivasta piirustuksessa?*	*Taitelija kuvasi nuorena mieluiten simmua piirustuksessa.*
						'eye'	'When the artist was young, did (s)he mostly depicted the sky in drawing?'	'When the artist was young, (s)he mostly depicted the eye in drawing.'
4	2	1	B	ESS	CVC.CV.CV	*silmuna*	*Miksi luulet, että isän lempitaitelija oli outo kaveri?*	*Taitelija kuvasi vaimonsa mieluiten silmuna piirustuksessa.*
						'as the bud of a leaf'	'Why do you think that father's favourite artist is a strange guy?'	'The artist mostly depicted his wife as the bud of a leaf in drawing.'
4	2	2	B	ESS	CVCg.CV.CV	*simmuna*	*Miksi luulet, että isän lempitaitelija oli outo kaveri?*	*Taitelija kuvasi vaimonsa mieluiten simmuna piirustuksessa.*
						'as an eye'	'Why do you think that father's favourite artist is a strange guy?'	'The artist mostly depicted his wife as an eye in drawing.'
4	2	3	N	ESS	CVC.CV.CV	*silmuna*	*Kuvasiko taitelija vaimonsa mieluiten kotkana piirustuksessa?*	*Taitelija kuvasi vaimonsa mieluiten silmuna piirustuksessa.*
						'as the bud of a leaf'	'Did the artist mostly depict his wife as an eagle in drawing?'	'The artist mostly depicted his wife as the bud of a leaf in drawing.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
4	2	1	N	ESS	CVCg.CV.CV	*simmuna*	*Kuvasiko taitelija vaimonsa mieluiten kotkana piirustuksessa?*	*Taitelija kuvasi vaimonsa mieluiten simmuna piirustuksessa.*
						'as an eye'	'Did the artist mostly depict his wife as an eagle in drawing?'	'The artist mostly depicted his wife as an eye in drawing.'
4	3	3	B	NOM	CVC.CV	*kalmo*	*Miksi äiti oli vihainen?*	*Lastenleffassa oli kalmo lopuksi.*
						'corpse'	'Why was mother angry?'	'In the children's movie, there was corpse at the end.'
4	3	1	B	NOM	CVCg.CV	*kammo*	*Miksi äiti oli vihainen?*	*Lastenleffasta tuli kammo lopuksi.*
						'phobia'	'Why was mother angry?'	'The children's movie caused a phobia at the end.'
4	3	2	N	NOM	CVC.CV	*kalmo*	*Miksi äiti valitteli? Oliko lastenleffassa ase lopuksi?*	*Lastenleffassa oli kalmo lopuksi.*
						'corpse'	'Why was mother angry? Was there a weapon in the children's movie at the end?'	'In the children's movie, there was corpse at the end.'
4	3	3	N	NOM	CVCg.CV	*kammo*	*Miksi äiti valitteli? Tuliko lastenleffasta suru lopuksi?*	*Lastenleffasta tuli kammo lopuksi.*
						'phobia'	'Why was mother angry? Did the children's movie cause sadness at the end?'	'The children's movie caused a phobia at the end.'
4	3	1	B	PRT	CVC.CVV	*kalmoa*	*Miksi äiti oli vihainen?*	*Lastenleffa esitteli kalmoa lopuksi.*
						'corpse'	'Why was mother angry?'	'The children's movie showed a corpse at the end.'
4	3	2	B	PRT	CVCg.CVV	*kammoa*	*Miksi äiti oli vihainen?*	*Lastenleffa käsitteli kammoa lopuksi.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'phobia'	'Why was mother angry?'	'The children's movie dealt with phobia at the end.'
4	3	3	N	PRT	CVC.CVV	*kalmoa*	*Miksi äiti oli vihainen? Esittelikö lastenleffa aavetta lopuksi?*	*Lastenleffa esitteli kalmoa lopuksi.*
						'corpse'	'Why was mother angry? Did the children's movie show a ghost at the end?'	'The children's movie showed a corpse at the end.'
4	3	1	N	PRT	CVCg.CVV	*kammoa*	*Miksi äiti oli vihainen? Käsittelikö lastenleffa surua lopuksi?*	*Lastenleffa käsitteli kammoa lopuksi.*
						'phobia'	'Why was mother angry? Did the children's movie deal with sadness at the end?'	'The children's movie dealt with phobia at the end.'
4	3	2	B	ESS	CVC.CV.CV	*kalmona*	*Onko hänellä iso rooli?*	*Ensi-illassa Sari näytteli kalmona lopuksi.*
						'as a corpse'	'Did she have a big role?'	'In the premiere, Sari played a corpse at the end.'
4	3	3	B	ESS	CVCg.CV.CV	*kammona*	*Oliko se uhkaava?*	*Oireita psykiatri käsitteli kammona lopuksi.*
						'as phobia'	'Was it threatening?'	'The psychiatrist described the symptoms as a phobia in the end.'
4	3	1	N	ESS	CVC.CV.CV	*kalmona*	*Näyttelikö Sari ensi-illassa aavena lopuksi?*	*Ensi-illassa Sari näytteli kalmona lopuksi.*
						'as a corpse'	'Did Sari play a ghost at the end?'	'In the premiere, Sari played a corpse at the end.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
4	3	2	N	ESS	CVCg.CV.CV	*kammona*	*Käsittelikö psykiatri oireita maniana lopuksi?*	*Oireita psykiatri käsitteli kammona lopuksi.*
						'as phobia'	'Did the psychiatrist describe the symptoms as a mania in the end.'	'The psychiatrist described the symptoms as a phobia in the end.'
4	4	1	B	NOM	CVC.CV	*Valma*	*Kuinka se meni, kun isoisä avasi vuosijuhlaa?*	*Kainalossaan hänellä oli Valma kaikkien nähtävillä.*
						'Valma (female name)'	'How did it go when the grandfather opened the anniversary?'	'On his arm was [lit. 'under his arm pit he had'] Valma for all to see.'
4	4	3	B	NOM	CVCg.CV	*vamma*	*Kuinka se meni, kun isoisä avasi vuosijuhlaa?*	*Kainalossaan hänellä oli vamma kaikkien nähtävillä.*
						'injury'	'How did it go when the grandfather opened the anniversary?'	'In his armpit he had an injury for all to see.'
4	4	2	N	NOM	CVC.CV	*Valma*	*Oliko hänellä kainalossaan Arja kaikkien nähtävillä?*	*Kainalossaan hänellä oli Valma kaikkien nähtävillä.*
						'Valma (female name)'	'Was Arja on his arm for all to see?'	'On his arm was [lit. 'under his arm pit he had'] Valma for all to see.'
4	4	1	N	NOM	CVCg.CV	*vamma*	*Oliko hänellä kainalossaan hikiläikkä kaikkien nähtävillä?*	*Kainalossaan hänellä oli vamma kaikkien nähtävillä.*
						'injury'	'Did he have a sweat stain in his armpit for all to see?'	'In his armpit he had an injury for all to see.'
4	4	3	B	PRT	CVC.CVV	*Valmaa*	*Kuinka se meni, kun isoisä avasi vuosijuhlaa?*	*Kainalossaan hän piteli Valmaa kaikkien nähtävillä.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'Valma (female name)'	'How did it go when the grandfather opened the anniversary?'	'He kept Valma on his arm [lit. 'under his armpit'] for all to see.'
4	4	2	B	PRT	CVCg.CVV	*vammaa*	*Kuinka se meni, kun isoisä avasi vuosijuhlaa?*	*Kainalossaan hän paranteli vammaa kaikkien nähtävillä.*
						'injury'	'How did it go when the grandfather opened the anniversary?'	'He was healing an injury in his armpit for all to see.'
4	4	1	N	PRT	CVC.CVV	*Valmaa*	*Pitelikö hän kainalossaan Arjaa kaikkien nähtävillä?*	*Kainalossaan hän piteli Valmaa kaikkien nähtävillä.*
						'Valma (female name)'	'Did he keep Arja on his arm [lit. 'under his armpit'] for all to see?'	'He kept Valma on his arm [lit. 'under his armpit'] for all to see.'
4	4	3	N	PRT	CVCg.CVV	*vammaa*	*Paranteliko hän kainalossaan tulehdusta kaikkien nähtävillä?*	*Kainalossaan hän paranteli vammaa kaikkien nähtävillä.*
						'injury'	'Was he healing an inflammation in his armpit for all to see?'	'He was healing an injury in his armpit for all to see.'
4	4	2	B	ESS	CVC.CV.CV	*Valmana*	*Mitä isoisä luuli?*	*Henkilöä hän piti Valmana kaikkien mielestä.*
						'as Valma (female name)'	'What did grandfather believe?'	'He took the person to be Valma, everyone agrees.'
4	4	1	B	ESS	CVCg.CV.CV	*vammana*	*Mitä isoisä ajatteli?*	*Heikkoutta hän piti vammana kaikkien mielestä.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'as an injury'	'What did grandfather think?'	'He took the weakness to be an injury, everyone agrees.'
4	4	3	N	ESS	CVC.CV.CV	*Valmana*	*Pitikö hän henkilöä Arjana kaikkien mielestä?*	*Henkilöä hän piti Valmana kaikkien mielestä.*
						'as Valma (female name)'	'Does everyone agree he took the person to be Arja?'	'He took the person to be Valma, everyone agrees.'
4	4	2	N	ESS	CVCg.CV.CV	*vammana*	*Pitikö hän heikkoutta hyveenä kaikkien mielestä?*	*Heikkoutta hän piti vammana kaikkien mielestä.*
						'as an injury'	'Does everyone agree he took the weakness to be a virtue?'	'He took the weakness to be an injury, everyone agrees.'
4	5	2	B	NOM	CVC.CV	*almu*	*Miksi Ville ei halunnut, mitä hänen veljensä tarjosi?*	*Ylpeälle jokainen apu on usein almu kuitenkin.*
						'alms'	'Why did Ville not want what his brother offered?'	'To a proud [person], every help is often alms anyhow.'
4	5	1	B	NOM	CVCg.CV	*ammu*	*Miksi Harri ei ymmärtänyt, mistä Suvi puhui?*	*Lapselle jokainen lehmä on usein ammu kuitenkin.*
						'moo-moo (baby talk for cow)'	'Why did Harri not understand what Suvi was talking about?'	'To a child, every cow is often a moo-moo anyhow.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
4	5	3	N	NOM	CVC.CV	*almu*	*Onko ylpeälle jokainen apu usein onni kuitenkin?*	*Ylpeälle jokainen apu on usein almu kuitenkin.*
						'alms'	'Is every help often a blessing to a proud [person] anyhow?'	'To a proud [person], every help is often alms anyhow.'
4	5	2	N	NOM	CVCg.CV	*ammu*	*Onko lapselle jokainen lehmä usein heppa kuitenkin?*	*Lapselle jokainen lehmä on usein ammu kuitenkin.*
						'moo-moo (baby talk for cow)'	'Is every cow often a horsie to a child anyhow?'	'To a child, every cow is often a moo-moo anyhow.'
4	5	1	B	PRT	CVC.CVV	*almua*	*Miksi Ville ei halunnut, mitä hänen veljensä tarjosi?*	*Ylpeä välttelee usein almua kuitenkin.*
						'alms'	'Why did Ville not want what his brother offered?'	'A proud [person] often avoids alms anyhow.'
4	5	3	B	PRT	CVCg.CVV	*ammua*	*Miksi Harri ei ymmärtänyt, mistä Suvi puhui?*	*Lapsi kutsuu usein ammua kuitenkin.*
						'moo-moo (baby talk for cow)'	'Why did Harri not understand what Suvi was talking about?'	'A/the child often calls moo-moo anyhow.'
4	5	2	N	PRT	CVC.CVV	*almua*	*Välttelee ylpeä usein työtä kuitenkin?*	*Ylpeä välttelee usein almua kuitenkin.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'alms'	'Does a proud [person] often avoids work anyhow?'	'A proud [person] often avoids alms anyhow.'
4	5	1	N	PRT	CVCg.CVV	*ammua*	*Kutsuuko lapsi usein heppaa kuitenkin?*	*Lapsi kutsuu usein ammua kuitenkin.*
						'moo-moo (baby talk for cow)'	'Does the child often call horsie anyhow?'	'The child often calls moo-moo anyhow.'
4	5	3	B	ESS	CVC.CV.CV	*almuna*	*Miksi Ville ei halunnut, mitä hänen veljensä tarjosi?*	*Ylpeä käsittää avun usein almuna kuitenkin.*
						'as alms'	'Why did Ville not want what his brother offered?'	'A proud [person] often takes help as alms anyhow.'
4	5	2	B	ESS	CVCg.CV.CV	*ammuna*	*Miksi Harri ei ymmärtänyt, mistä Suvi puhui?*	*Lapsi käsittää lehmän usein ammuna kuitenkin.*
						'as a moo-moo (baby talk for cow)'	'Why did Harri not understand what Suvi was talking about?'	'A child often describesa cow as moo-moo anyhow.'
4	5	1	N	ESS	CVC.CV.CV	*almuna*	*Käsittääkö ylpeä avun usein onnena kuitenkin?*	*Ylpeä käsittää avun usein almuna kuitenkin.*
						'as alms'	'Does a proud [person] often describe help as a blessing anyhow?'	'A proud [person] often describes help as alms anyhow.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
4	5	3	N	ESS	CVCg.CV.CV	*ammuna*	*Käsittääkö lapsi lehmän usein heppana kuitenkin?*	*Lapsi käsittää lehmän usein ammuna kuitenkin.*
						'as a moo-moo moo (baby talk for cow)'	'Does a child often describes a cow as horsie anyhow?'	'A child often describesa cow as moo-moo anyhow.'
4	6	3	B	NOM	CVC.CV	*nurmi*	*Miksi pidit viimeisestä talostanne?*	*Meillä oli tontin takana nurmi kukkineen.*
						'grassland, lawn'	'Why did you like your previous house?'	'Behind [our] plot of land we had grassland with flowers.'
4	6	2	B	NOM	CVCg.CV	*nummi*	*Miksi pidit viimeisestä talostanne?*	*Meillä oli tontin takana nummi kukkineen.*
						'moor'	'Why did you like your previous house?'	'Behind [our] plot of land we had a moor with flowers.'
4	6	1	N	NOM	CVC.CV	*nurmi*	*Oliko teillä tontin takana metsä kukkineen?*	*Meillä oli tontin takana nurmi kukkineen.*
						'grassland, lawn'	'Did you have a forest with flowers behind [your] plot of land?'	'Behind [our] plot of land we had grassland with flowers.'
4	6	3	N	NOM	CVCg.CV	*nummi*	*Oliko teillä tontin takana metsä kukkineen?*	*Meillä oli tontin takana nummi kukkineen.*
						'moor'	'Did you have a forest with flowers behind [your] plot of land?'	'Behind [our] plot of land we had a moor with flowers.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
4	6	2	B	PRT	CVC.CVV	*nurmia*	*Miksi ostit edellisen talosi?*	*Halusin ihailla tontin takana nurmia kukkineen.*
						'grassland, lawn'	'Why did you buy your previous house?'	'I wanted to admire the grassland with flowers behind the plot of land.'
4	6	1	B	PRT	CVCg.CVV	*nummia*	*Miksi ostit edellisen talosi?*	*Halusin ihailla tontin takana nummia kukkineen.*
						'moor'	'Why did you buy your previous house?'	'I wanted to admire the moor with flowers behind the plot of land.'
4	6	3	N	PRT	CVC.CVV	*nurmia*	*Halusitko ihailla tontin takana metsää kukkineen?*	*Halusin ihailla tontin takana nurmia kukkineen.*
						'grassland, lawn'	'Did you want to admire the forest with flowers behind the plot of land?'	'I wanted to admire the grassland with flowers behind the plot of land.'
4	6	2	N	PRT	CVCg.CVV	*nummia*	*Halusitko ihailla tontin takana metsää kukkineen?*	*Halusin ihailla tontin takana nummia kukkineen.*
						'moor'	'Did you want to admire the forest with flowers behind the plot of land?'	'I wanted to admire the moor with flowers behind the plot of land.'
4	6	1	B	ESS	CVC.CV.CV	*nurmina*	*Miksi kaupunginjohtaja riitelee vihreiden kanssa?*	*Ne haluavat rauhoittaa alueita tontin takana nurmina kukkineen.*
						'as grassland, lawn'	'Why does the mayor quarrel with the greens?'	'They want to protect the areas behind the plot of land as a grassland with flowers.'
4	6	3	B	ESS	CVCg.CV.CV	*nummina*	*Miksi kaupunginjohtaja riitelee vihreiden kanssa?*	*Ne haluavat rauhoittaa alueita tontin takana nummina kukkineen.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						‘as grass-land, lawn’	‘Why does the mayor quarrel with the greens?’	‘They want to protect the areas behind the plot of land as a moor with flowers.’
4	6	2	N	ESS	CVC.CV.CV	*nurmina*	*Haluavatko vihreät rauhoittaa alueita tontin takana puistona kukkineen?*	*Ne haluavat rauhoittaa alueita tontin takana nurmina kukkineen.*
						‘as a moor’	‘Do the greens want to protect the areas behind the plot of land as a forest with flowers?’	‘They want to protect the areas behind the plot of land as a grassland with flowers.’
4	6	1	N	ESS	CVCg.CV.CV	*nummina*	*Haluavatko vihreät rauhoittaa alueita tontin takana puistona kukkineen?*	*Ne haluavat rauhoittaa alueita tontin takana nummina kukkineen.*
						‘as a moor’	‘Do the greens want to protect the areas behind the plot of land as a forest with flowers?’	‘They want to protect the areas behind the plot of land as a moor with flowers.’
5	1	1	B	NOM	CVC.CV	*kulma*	*Mitä Simo valitteli?*	*Hänen mielestä paperissa oli kulma tiistaina.*
						‘corner’	‘[About] what did Simo complain?’	‘In his opinion the paper had a fold [lit. ‘there was an edge in the paper’] on Tuesday.’
5	1	2	B	NOM	CVV.CV	*kuuma*	*Mitä Simo väitteli?*	*Hänen mielestä saunassa oli kuuma tiistaina.*
						‘hot’	‘[About] what did Simo complain?’	‘In his opinion it was hot in sauna on Tuesday.’

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
5	1	3	N	NOM	CVC.CV	*kulma*	*Oliko Simon mielestä paperissa läikkä tiistaina?*	*Hänen mielestä paperissa oli kulma tiistaina.*
						'corner'	'Did the paper have an spot [lit. 'there was a spot in the paper'] in Simo's opinion on Tuesday?'	'In his opinion the paper had a fold [lit. 'there was an edge in the paper'] on Tuesday.'
5	1	1	N	NOM	CVV.CV	*kuuma*	*Oliko Simon mielestä saunassa kylmä tiistaina?*	*Hänen mielestä saunassa oli kuuma tiistaina.*
						'hot'	'Was it in Simo's opinion cold in sauna on Tuesday?'	'In his opinion it was hot in sauna on Tuesday.'
5	1	2	B	PRT	CVC.CVV	*kulmaa*	*Mitä näkyi ikkunasta?*	*Kadulla Jussi kierteli kulmaa tiistaina.*
						'corner'	'What could you see through the window?'	'On the street, Jussi walked around the corner on Tuesday.'
5	1	3	B	PRT	CVV.CVV	*kuumaa*	*Mitä näkyi ikkunasta?*	*Talossa Jussi vältteli kuumaa tiistaina.*
						'hot'	'What could you see through the window?'	'In the house, Jussi complained about the heat on Tuesday.'
5	1	1	N	PRT	CVC.CVV	*kulmaa*	*Kiertelikö Jussi kadulla autoa tiistaina?*	*Kadulla Jussi kierteli kulmaa tiistaina.*
						'corner'	'Did Jussi walk around a car on the street on Tuesday?'	'On the street, Jussi walked around the corner on Tuesday.'
5	1	2	N	PRT	CVV.CVV	*kuumaa*	*Välttelikö Jussi talossa melua tiistaina?*	*Talossa Jussi vältteli kuumaa tiistaina.*
						'hot'	'Did Jussi complain about the noise in the house on Tuesda?.'	'In the house, Jussi complained about the heat on Tuesday.'
5	1	3	B	ESS	CVC.CV.CV	*Kulmana*	*Mitä naapuri kertoi?*	*Uusi kahvila aukaisi Kulmana tiistaina.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'as Kulma ('the corner')'	'What did the neighbour tell [you]?'	'The new cafe opened under the name [lit. 'as'] the Corner on Tuesday.'
5	1	1	B	ESS	CVV.CV.CV	*kuumana*	*Miksi mummo oli mielissään?*	*Hellalla pysyi velli kuumana tiistaina.*
						'being hot'	'Why was grandma pleased?'	'The gruel remained warm on the stove on Tuesday.'
5	1	2	N	ESS	CVC.CV.CV	*Kulmana*	*Aukaisiko uusi kahvila Paussina tiistaina?*	*Uusi kahvila aukaisi Kulmana tiistaina.*
						'as Kulma ('the corner')'	'Did the new cafe open under the name [lit. 'as'] the Pause on Tuesday?'	'The new cafe opened under the name [lit. 'as'] the Corner on Tuesday.'
5	1	3	N	ESS	CVV.CV.CV	*kuumana*	*Pysyikö hellalla velli kylmänä tiistaina?*	*Hellalla pysyi velli kuumana tiistaina.*
						'being hot'	'Did the porridge remain cold on the stove on Tuesday?'	'The porridge remained warm on the stove on Tuesday.'
5	2	2	B	NOM	CVC.CV	*talja*	*Mitä muistat naapurin Tommista?*	*Hänen rekensä penkillä oli talja pari vuotta.*
						'fleeze, animal hide'	'What do you remember about the neighbour's [son] Tommi?'	'On the seat of his sleigh was a fleece for a few years.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
5	2	3	B	NOM	CVV.CV	*taaja*	*Mitä lehdessä luki?*	*Porin jatseilla yleisö oli taaja pari vuotta.*
						'dense'	'What does it say in the newspaper?'	'The audience was packed [lit. 'dense'] at the Pori jazz [festival] for a few years.'
5	2	1	N	NOM	CVC.CV	*talja*	*Oliko Tommin reen penkillä peitto pari vuotta?*	*Hänen rekensä penkillä oli talja pari vuotta.*
						'fleeze, animal hide'	'Was there a cover on the seat of Tommi's sleigh for a few years?'	'On the seat of his sleigh was a fleece for a few years.'
5	2	2	N	NOM	CVV.CV	*taaja*	*Oliko Porin jatseilla yleisö harva pari vuotta?*	*Porin jatseilla yleisö oli taaja pari vuotta.*
						'dense'	'Was the audience sparse at the Pori jazz [festival] for a few years?'	'The audience was packed [lit. 'dense'] at the Pori jazz [festival] for a few years.'
5	2	3	B	PRT	CVC.CVV	*taljaa*	*Mitä muistat naapurin Tommista?*	*Rekensä penkillä hän piti taljaa pari vuotta.*
						'fleeze, animal hide'	'What do you remember about the neighbour's [son] Tommi?'	'On the seat of his sleigh he kept a fleece for a few years.'
5	2	1	B	PRT	CVV.CVV	*taajaa*	*Kuinka Tommi selvisi yksinäisyydestä?*	*Hänen leffassa käyntinsä oli taajaa pari vuotta.*
						'dense'	'How did Tommi deal with [his] loneliness?'	'His visits to the movies were frequent [lit. 'dense'] for a few years.'
5	2	2	N	PRT	CVC.CVV	*taljaa*	*Pitikö hän rekensä penkillä peittoa pari vuotta?*	*Rekensä penkillä hän piti taljaa pari vuotta.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						‘fleeze, animal hide’	‘Did he keep a cover on the seat of his sleigh for a few years?’	‘On the seat of his sleigh he kept a fleece for a few years.’
5	2	3	N	PRT	CVV.CVV	*taajaa*	*Oliko hänen leffassakäyntinsä harvaa pari vuotta?*	*Hänen leffassakäyntinsä oli taajaa pari vuotta.*
						‘dense’	‘Were his visits to the movies sparse for a few years?’	‘His visits to the movies were frequent [lit. ‘dense’] for a few years.’
5	2	1	B	ESS	CVC.CV.CV	*taljana*	*Mitä muistat naapurin Tommista?*	*Tommin reessä viltti toimi taljana pari vuotta.*
						‘as a fleeze, animal hide’	‘What do you remember about the neighbour’s [son] Tommi?’	‘In Tommi’s sleigh, a blanket served as a fleece for a few years.’
5	2	2	B	ESS	CVV.CV.CV	*taajana*	*Kuinka Tommi selvisi yksinäisyydestä?*	*Hänen leffassakäyntinsä jatkui taajana pari vuotta.*
						‘being dense’	‘How did Tommi deal with [his] loneliness?’	‘His visits to the movies continued frequently [lit. ‘as dense’] for a few years.’
5	2	3	N	ESS	CVC.CV.CV	*taljana*	*Toimiko viltti Tommin reessä kattona pari vuotta?*	*Tommin reessä viltti toimi taljana pari vuotta.*
						‘as a fleeze, animal hide’	‘Did a blanket serve as a roof in Tommi’s sleigh for a few years?’	‘In Tommi’s sleigh, a blanket served as a fleeze for a few years.’

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
5	2	1	N	ESS	CVV.CV.CV	*taajana*	*Jatkuiko hänen leffassakäyntinsä harvana pari vuotta?*	*Hänen leffassakäyntinsä jatkui taajana pari vuotta.*
						‘being dense’	‘Did his visits to the movies continue sparsely for a few years?’	‘His visits to the movies continued frequently [lit. ‘as dense’] for a few years.’
5	3	3	B	NOM	CVC.CV	*Vanja*	*Mitä Suvi kertoi menneisyydestä?*	‘Suvi’s best friend was called Vanjain school times [lit. ‘being school aged’].’
						‘Vanja (male name)’	‘What did Suvi tell about the past?’	‘In school times [lit. ‘being school aged’] Suvi’s best friend was called Vanja.’
5	3	1	B	NOM	CVV.CV	*vaaja*	*Mitä mestari painotti?*	*Tärkein esine oli nimeltään vaaja koulukirjassa.*
						‘wedge’	‘What [point] did the foreman stress?’	‘The most important object was called a wedge in the school book.’
5	3	2	N	NOM	CVC.CV	*Vanja*	*Oliko Suvin paras kaveri nimeltään Aleksi kouluikäisenä?*	*Suvin paras kaveri oli nimeltään Vanja kouluikäisenä.*
						‘Vanja (male name)’	‘Was Suvi’s best friend was called Aleksi In school times [lit. ‘being school aged’]?’	‘Suvi’s best friend was called Vanjain school times [lit. ‘being school aged’].’
5	3	3	N	NOM	CVV.CV	*vaaja*	*Onko työprosessin tärkein esine nimeltään taltta koulukirjassa?*	*Tärkein esine on nimeltään vaaja koulukirjassa.*
						‘wedge’	‘Was the most important object of the work process called a chisel in the school book?’	‘The most important object was called a wedge in the school book.’

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
5	3	1	B	PRT	CVC.CVV	*Vanjaa*	*Mitä melu oli?*	*Kaverit huutelivat tänään Vanjaa koulutiellä.*
						'Vanja (male name)'	'What is [that] noise?'	'[His] friends called Vanja on the way to school today.'
5	3	2	B	PRT	CVV.CVV	*vaajaa*	*Mikä oli ongelma?*	*Työmiehet tarvitsivat tänään vaajaa koulutuksessa.*
						'wedge'	'What was the problem?'	'The workmen needed a wedge for [lit. 'in'] the training today.'
5	3	3	N	PRT	CVC.CVV	*Vanjaa*	*Huutelivatko kaverit tänään Aleksia koulutiellä?*	*Kaverit huutelivat tänään Vanjaa koulutiellä.*
						'Vanja (male name)'	'Did [his] friends call Aleksi on the way to school today?'	'[His] friends called Vanja on the way to school today.'
5	3	1	N	PRT	CVV.CVV	*vaajaa*	*Tarvitsivatko työmiehet tänään lekaa koulutuksessa?*	*Työmiehet tarvitsivat tänään vaajaa koulutuksessa.*
						'wedge'	'Did the workmen need a sledgehammer for [lit. 'in'] the training today?'	'The workmen needed a wedge for [lit. 'in'] the training today.'
5	3	2	B	ESS	CVC.CV.CV	*Vanjana*	*Miksi serkku on huvittunut?*	*Setä esiintyi tänään Vanjana koulunäytelmässä.*
						'as Vanja (male name)'	'Why is the cousin amused?'	'The uncle appeared as Vanja in the school play today.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
5	3	3	B	ESS	CVV.CV.CV	*vaajana*	*Miksi mestari on ärsyyntynyt?*	*Puupalaa käytettiin tänään vaajana koulutuksessa.*
						'as a wedge'	'Why is the foreman displeased?'	'A piece of wood was used as a wedge in the training today.'
5	3	1	N	ESS	CVC.CV.CV	*Vanjana*	*Esiintyikö setä tänään Otellona koulunäytelmässä?*	*Setä esiintyi tänään Vanjana koulunäytelmässä.*
						'as Vanja (male name)'	'Did the uncle appeared as Othello in the school play today?'	'The uncle appeared as Vanja in the school play today.'
5	3	2	N	ESS	CVV.CV.CV	*vaajana*	*Käytettiinkö puupalaa tänään talttana koulutuksessa?*	*Puupalaa käytettiin tänään vaajana koulutuksessa.*
						'as a wedge'	'Was a piece of wood used as a chisel in the training today?'	'A piece of wood was used as a wedge in the training today.'
5	4	1	B	NOM	CVC.CV	*halva*	*Miksi koira oli hermostunut?*	*Perheen nuorella tytöllä oli kädessään halva näkyvästi.*
						'halva'	'Why was the dog nervous?'	'The young daughter of the family visibly had halva in her hand.'
5	4	3	B	NOM	CVV.CV	*haava*	*Miksi koira oli hermostunut?*	*Perheen nuorella tytöllä oli kädessään haava näkyvästi.*
						'wound'	'Why was the dog nervous?'	'The young daughter of the family visibly had a wound in her hand.'
5	4	2	N	NOM	CVC.CV	*halva*	*Oliko perheen nuorella tytöllä kädessään nakki näkyvästi?*	*Perheen nuorella tytöllä oli kädessään halva näkyvästi.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'halva'	'Did the young daughter of the family visibly have a sausage in her hand?'	'The young daughter of the family visibly had halva in her hand.'
5	4	1	N	NOM	CVV.CV	*haava*	*Oliko perheen nuorella tytöllä kädessään tulehdus näkyvästi?*	*Perheen nuorella tytöllä oli kädessään haava näkyvästi.*
						'wound'	'Did the young daughter of the family visibly have an inflammation in her hand?'	'The young daughter of the family visibly had a wound in her hand.'
5	4	3	B	PRT	CVC.CVV	*halvaa*	*Miksi koira oli hermostunut?*	*Perheen nuorin tyttö piteli kädessään halvaa näkyvästi.*
						'halva'	'Why was the dog nervous?'	'The young daughter of the family visibly held halva in her hand.'
5	4	2	B	PRT	CVV.CVV	*haavaa*	*Miksi he menevät lääkärille?*	*Perheen nuorin tyttö paranteli kädessään haavaa näkyvästi.*
						'wound'	'Why did they go to the doctor?'	'The young daughter of the family visibly healed wound in her hand.'
5	4	1	N	PRT	CVC.CVV	*halvaa*	*Pitelikö perheen nuorin tyttö kädessään suklaata näkyvästi?*	*Perheen nuorin tyttö piteli kädessään halvaa näkyvästi.*
						'halva'	'Did the young daughter of the family visibly hold chocolate in her hand?'	'The young daughter of the family visibly held halva in her hand.'
5	4	3	N	PRT	CVV.CVV	*haavaa*	*Paranteliko perheen nuorin tyttö kädessään tulehdusta näkyvästi?*	*Perheen nuorin tyttö paranteli kädessään haavaa näkyvästi.*
						'wound'	'Did the young daughter of the family visibly heal an inflammation in her hand?'	'The young daughter of the family visibly healed wound in her hand.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
5	4	2	B	ESS	CVC.CV.CV	*halvana*	*Mikä on uusin kohu-uutinen?*	*Kaupungilla tahnaa myytiin halvana näkyvästi.*
						'as halva'	'What is the latest sensational news story¿'	'In the city they visibly sold paste as halva.'
5	4	1	B	ESS	CVV.CV.CV	*haavana*	*Mikä on uusin kohu-uutinen?*	*Sairaalassa vammaa paranneltiin haavana näkyvästi.*
						'as a wound'	'What is the latest sensational news story¿'	'In the hospital they visibly cured an injury as a wound.'
5	4	3	N	ESS	CVC.CV.CV	*halvana*	*Myytiinkö kaupungilla tahnaa maapähkinävoina näkyvästi?*	*Kaupungilla tahnaa myytiin halvana näkyvästi.*
						'as halva'	'Did they visibly sell paste as peanut butter in the city?'	'In the city they visibly sold paste as halva.'
5	4	2	N	ESS	CVV.CV.CV	*haavana*	*Paranneltiinko sairaalassa vammaa tulehduksena näkyvästi?*	*Sairaalassa vammaa paranneltiin haavana näkyvästi.*
						'as a wound'	'Did they visibly cure an injury as a inflammation in the hospital?'	'In the hospital they visibly cured an injury as a wound.'
5	5	2	B	NOM	CVC.CV	*Talvi*	*Mitä uutta oli tänään jalkapalloharjoituksissa?*	*Meidän ryhmään tuli tyttö nimeltään Talvi sattumalta.*
						'Talvi (female name)'	'What news were there today in the football practice?'	'A girl named Talvi came into our group accidentally.'
5	5	1	B	NOM	CVV.CV	*Taavi*	*Mitä uutta oli tänään jalkapalloharjoituksissa?*	*Meidän ryhmään tuli poika nimeltään Taavi sattumalta.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'Taavi (male name)'	'What news were there today in the football practice?'	'A boy named Taavi came into our group accidentally.'
5	5	3	N	NOM	CVC.CV	*Talvi*	*Tuliko teidän ryhmään tyttö nimeltään Meri sattumalta?*	*Meidän ryhmään tuli tyttö nimeltään Talvi sattumalta.*
						'Talvi (female name)'	'Did a girl named Meri come into your group accidentally?'	'A girl named Talvi came into our group accidentally.'
5	5	2	N	NOM	CVV.CV	*Taavi*	*Tuliko teidän ryhmään poika nimeltään Erkki sattumalta?*	*Meidän ryhmään tuli poika nimeltään Taavi sattumalta.*
						'Taavi (male name)'	'Did a boy named Erkki come into your group accidentally?'	'A boy named Taavi came into our group accidentally.'
5	5	1	B	PRT	CVC.CVV	*Talvia*	*Mitä kaiuttimissa sanottiin?*	*Portille kahdeksan kutsuttiin Talvia sattumalta.*
						'Talvi (female name)'	'What was the loud speaker announcement [lit. 'what is said in the speakers']?'	'They summoned Talvi to gate twelve accidentally.'
5	5	3	B	PRT	CVV.CVV	*Taavia*	*Mitä kaiuttimissa sanottiin?*	*Portille kahdeksan kutsuttiin Taavia sattumalta.*
						'Taavi (male name)'	'What was the loud speaker announcement [lit. 'what is said in the speakers']?'	'They summoned Taavi to gate twelve accidentally.'

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
5	5	2	N	PRT	CVC.CVV	*Talvia*	*Kutsuttiinko portille kahdeksan Meriä sattumalta?*	*Portille kahdeksan kutsuttiin Talvia sattumalta.*
						‘Talvi (female name)’	‘Did they summon Meri to gate twelve accidentally?’	‘They summoned Talvi to gate twelve accidentally.’
5	5	1	N	PRT	CVV.CVV	*Taavia*	*Kutsuttiinko portille kahdeksan Erkkiä sattumalta?*	*Portille kahdeksan kutsuttiin Taavia sattumalta.*
						‘Taavi (male name)’	‘Did they summon Erkki to gate twelve accidentally?’	‘They summoned Taavi to gate twelve accidentally.’
5	5	3	B	ESS	CVC.CV.CV	*Talvina*	*Mistä tuli hämmennys?*	*Ryhmän uutta tyttöä pidettiin Talvina sattumalta.*
						‘as Talvi (female name)’	‘Where did the confusion come from?’	‘The new girl in the group was [mis]taken for Talvi accidentally.’
5	5	2	B	ESS	CVV.CV.CV	*Taavina*	*Mistä tuli hämmennys?*	*Ryhmän uutta poikaa pidettiin Taavina sattumalta.*[5]
						‘as Taavi (male name)’	‘Where did the confusion come from?’	‘The new boy in the group was [mis]taken for Taavi accidentally.’

Continued on next page

[5]For participants s02 and s05, the frame sentence included the word *tyttöä* ‘girl’ instead of *poikaa* ‘boy’, but this only minimally affected plausibility and did not affect focus structure at all. Their materials were thus retained.

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
5	5	1	N	ESS	CVC.CV.CV	*Talvina*	*Pidettiinkö ryhmän uutta tyttöä Merinä sattumalta?*	*Ryhmän uutta tyttöä pidettiin Talvina sattumalta.*
						'as Talvi (female name)'	'Was the new girl in the group [mis]taken for Meri accidentally.'	'The new girl in the group was [mis]taken for Talvi accidentally.'
5	5	3	N	ESS	CVV.CV.CV	*Taavina*	*Pidettiinkö ryhmän uutta poikaa Erkkinä sattumalta?*	*Ryhmän uutta poikaa pidettiin Taavina sattumalta.*[6]
						'as Taavi (male name)'	'Was the new boy in the group [mis]taken for Erkkii accidentally.'	'The new boy in the group was [mis]taken for Taavi accidentally.'
5	6	3	B	NOM	CVC.CV	*palvi*	*Miksi tykkäsit työpaikasta?*	*Huoneessa tuoksui paljon palvi koko ajan.*
						'cured meat'	'Why did you like [your] workplace?'	'The room [lit. 'in the room'] smelled a lot like cured meat all the time.'
5	6	2	B	NOM	CVV.CV	*paavi*	*Miksi tykkäsit lomasta?*	*Roomassa puhui paljon paavi koko ajan.*
						'pope'	'Why did you like [your] holiday?'	'In Rome the pope spoke a lot all the time.'
5	6	1	N	NOM	CVC.CV	*palvi*	*Tuoksuiko huoneessa paljon kaakao koko ajan?*	*Huoneessa tuoksui paljon palvi koko ajan.*

Continued on next page

[6]For participants s03 and s06, the frame sentence included the word *tyttöä* 'girl' instead of *poikaa* 'boy'. Their responses were nevertheless not excluded from the analysis.

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'cured meat'	'Did the room [lit. 'in the room'] smell a lot like cocoa all the time?'	'The room [lit. 'in the room'] smelled a lot like cured meat all the time.'
5	6	3	N	NOM	CVV.CV	*paavi*	*Puhuiko Roomassa paljon Berluskooni koko ajan?*	*Roomassa puhui paljon paavi koko ajan.*
						'pope'	'Did Berlusconi speak a lot in Rome all the time?'	'In Rome the pope spoke a lot all the time.'
5	6	2	B	PRT	CVC.CVV	*palvia*	*Mitä oli reissulla parasta?*	*Roomassa syötiin paljon palvia koko ajan.*
						'cured meat'	'What was the best about the journey?'	'In Rome we ate a lot of cured meat all the time.'
5	6	1	B	PRT	CVV.CVV	*paavia*	*Mitä oli reissulla parasta?*	*Roomassa kuunneltiin paljon paavia koko ajan.*
						'pope'	'What was the best about the journey?'	'In Rome we heard the pope a lot all the time?'
5	6	3	N	PRT	CVC.CVV	*palvia*	*Syötiinkö Roomassa paljon juustoa koko ajan?*	*Roomassa syötiin paljon palvia koko ajan.*
						'cured meat'	'Did you we eat a lot of cheese in Rome all the time?'	'In Rome we ate a lot of cured meat all the time.'
5	6	2	N	PRT	CVV.CVV	*paavia*	*Kuunneltiinko Roomassa paljon Berluskoonia koko ajan?*	*Roomassa kuunneltiin paljon paavia koko ajan.*
						'pope'	'Did you hear Berlusconi in Rome a lot in Rome all the time?'	'In Rome we heard the pope a lot all the time?'
5	6	1	B	ESS	CVC.CV.CV	*palvina*	*Mitä piti aina korjata?*	*Lihaa pidettiin usein palvina koko ajan.*

Continued on next page

Table H.1 – *Continued from previous page*

Set	Pair	List	Foc.	Case	Syll. Struc.	Target	Context Question	Whole response
						'as cured meat'	'What did one always need to correct?'	'Meat was often [mis]taken for cured meat all the time.'
5	6	3	B	ESS	CVV.CV.CV	*paavina*	*Mitä piti aina korjata?*	*Pappia pidettiin usein paavina koko ajan.*
						'as the pope'	'What did one always need to correct?'	'The priest was often [mis]taken for the pope all the time.'
5	6	2	N	ESS	CVC.CV.CV	*palvina*	*Pidettiinkö lihaa usein paistina koko ajan?*	*Lihaa pidettiin usein palvina koko ajan.*
						'as cured meat'	'Was meat was often [mis]taken for a roast all the time?'	'Meat was often [mis]taken for cured meat all the time.'
5	6	1	N	ESS	CVV.CV.CV	*paavina*	*Pidettiinkö pappia usein filmitähtenä koko ajan?*	*Pappia pidettiin usein paavina koko ajan.*
						'as the pope'	'Was the priest was often [mis]taken for a film star all the time?'	'The priest was often [mis]taken for the pope all the time.'

Appendix I

Background information on participants of Production Experiment II

This appendix provides information about the subjects participating in the production experiment discussed in Chapter 4.

Table I.1: Age, sex, native language and dialect of participants of Production Experiment II.

Id	List	Age	Sex	Native language	Dialect
s01	1	23	f	Finnish	standard
s02	2	22	f	Finnish	standard
s03	3	21	f	Finnish	standard
s04	1	26	f	Finnish	standard
s05	1	41	f	Finnish	standard
s06	1	21	f	Finnish	standard and Helsinki dialect
s07	1	28	f	Finnish	standard
s08	1	22	f	Finnish	standard
s09	1	23	f	Finnish	something between standard and dialect
s10	1	20	f	Finnish	standard
s11	1	19	f	Finnish	standard
s12	1	28	m	Finnish	standard
s13	1	21	f	Finnish	standard
s14	1	28	f	Finnish	standard
s15	1	21	f	Finnish	standard
s16	1	25	f	Finnish	standard
s17	1	21	f	Finnish	standard
s18	1	21	f	Finnish	standard
s19	1	25	m	Finnish	standard
s20	1	21	f	Finnish, mother spoke both Finnish and English to her when she was younger than a year, then only Finnish	standard
s21	1	24	m	Finnish	standard
s22	1	29	f	Finnish	standard / weakly Oulu dialect
s23	1	22	f	Finnish	standard
s24	1	28	f	Finnish	standard
s25	1	29	f	Finnish	standard

Table I.2: Places of residency for participants of Production Experiment II.

Id	Lived in
s01	Vantaa, Tuusula, Järvenpää
s02	Helsinki
s03	Vantaa
s04	Helsinki
s05	Vantaa, Helsinki
s06	Helsinki ja Belgia (one year, frankophone area)
s07	Järvenpää, Helsinki, Göteborg (15 months), Espoo, Vantaa
s08	Helsinki
s09	Kiukainen (close to Pori), Kuopio, from school age on Helsinki
s10	Helsinki, Thessaloniki (Greece), Nuuk (Greenland)
s11	Kerava
s12	Helsinki
s13	Helsinki, Lauttasaari, Marjaniemi, Huopalahti, Klaukkala
s14	Vihti, Nummela, Helsinki, Göteborg (three years)
s15	Helsinki
s16	Helsinki
s17	Helsinki
s18	Espoo, Helsinki
s19	Espoo, Helsinki
s20	Tammisaari, Lohja, Helsinki
s21	Hollola, Kuortane, Helsinki
s22	Oulu, Helsinki (only after school)
s23	Helsinki
s24	Vantaa, Helsinki
s25	Espoo, Helsinki

Table I.3: Foreign languages spoken by participants of Production Experiment II.

Id	Other languages spoken
s01	English, Swedish, Spanish
s02	Swedish, English, French
s03	English, Swedish, French
s04	English, Swedish
s05	English, Swedish, German, French, Dutch, Spanish, Italian, Latin, Russian
s06	French, English, Swedish
s07	Swedish, English
s08	Swedish, English, Spanish, German, Danish
s09	French, English, Swedish, Spanish
s10	English, Swedish, French, Latin, Modern Greek, Italian, Greenlandic, Northern Saami
s11	English, Swedish
s12	Swedish, English, French, German, Russian, Italian, Croatian
s13	English, German, Swedish, French
s14	Swedish, English, German
s15	English, Swedish
s16	Swedish, English
s17	Swedish, German, English, Icelandic, Saami, Norwegian, Japanese
s18	English, Swedish, German
s19	English, Japanese
s20	English, French, Swedish, Tagalog
s21	English, Swedish
s22	English, Swedish, French
s23	Swedish, English, a little French, Russian and German
s24	French, English, Finnish sign language, Swedish, Spanish, Italian
s25	English, Swedish, Spanish, French

Table I.4: Information about possible speech pathologies and factors affecting speech physiology for participants of Production Experiment II.

Id	Speech pathology	Cold	Smoking
s01	no	no	no
s02	no	no	no
s03	no	no	no
s04	no	no	no
s05	no	no	no
s06	no	no	occasionally
s07	yes, slow reader, but did not affect experiment because sentences were short	no	no
s08	r problem as a child of less than 8 years	recently	no
s09	no	no	no
s10	no	no	no
s11	no	no	no
s12	no	no	stopped
s13	no	recently	occasionally
s14	no	no	no
s15	no	recently	stopped
s16	no	no	no
s17	no	recently	no
s18	no	recently	no
s19	no	no	no
s20	no	no	no
s21	no	no	no
s22	no	yes	no
s23	no	recently	no
s24	problems with hoarseness and loss of voice a few years ago	no	no
s25	no	no	no

Appendix J

Models of peak alignment for subsets of data from Production Experiment II

This appendix presents linear mixed-effects models of peak alignment for subsets of data from Production Experiment II (cf. Chapter 4). Section J.1 compares peak alignment relative to the beginning of the first syllable nucleus in disyllabic vs. trisyllabic items. Section J.2 presents separate models for each of the five first syllable contrast subsets (see Appendix H for the list of items). For each subset, it evaluates peak alignment relative to three segmental landmarks—the beginning of the first syllable nucleus, the end of the first mora and the beginning of the word.

Alignment was always measured as the absolute distance of the peak H from the respective landmark by subtracting the time of the landmark from the time of the peak.

J.1 Peak alignment in disyllabic and trisyllabic words

Table J.1 presents the best model of peak distance from the beginning of the nucleus in trisyllabic words (essive case; N = 881). The model showed significant effects of first syllable structure, with peaks of CVV.X and CVVC.X words being significantly later than those in CV.X words. Additionally, peaks were significantly later in narrow focus condition.

In contrast, the corresponding model for disyllabic words (nominative and partitive case; N = 1721) was considerably more complex (see Table J.2). It included only one significant main effect of initial syllable, indicating later peaks for words starting with CVC_g syllables. However, this main effect showed several interactions with other factors and overall the main effects and significant interactions rather seemed to cancel each other out. Since the number of items with initial CVC_g syllables was relatively small and more

Table J.1: Coefficients of the best linear mixed-effects model of the distance of the peak from the beginning of the first syllable nucleus (in ms) for trisyllabic words, with random by-item and by-subject slopes.

	Estimate	Std. Error	t value
(Intercept)	55.0607	8.3395	6.6024
1st syllable CVC	−9.5307	8.8574	−1.0760
1st syllable CVC_g	−13.8253	10.8121	−1.2787
1st syllable CVV	−26.8684	9.3461	−2.8748
1st syllable CVVC	−26.5081	8.9103	−2.9750
Narrow focus	11.0156	2.7036	4.0744

prone to influences from outliers than the other groups, it is likely that the model in Table J.2 is an overfit model.

J.2 Peak alignment in first syllable contrast subsets

The model in Table J.3 analysed peak alignment relative to the beginning of the first syllable nucleus for the data from contrast set 1 (first syllable CV vs. CVVC, $N = 516$). It indicated that the differences between CV and CVVC first syllables were significant also for this subset containing only real minimal item pairs. Peak alignment was thus significantly earlier for CVVC.X words than for CV.X words. The model also established a significant influence of focus condition and number of syllables, with later peaks in narrow focus and for trisyllabic words, while second syllable quantity did not appear to have a significant effect.

For the same subset, the best model relating the timing of the peak to the end of the first mora likewise confirmed the difference between the first syllable types CV and CVVC and the effect of number of syllables, but was not improved by including the factor focus (cf. Table J.4). For this measure, there was thus no indication of an effect of either focus condition or second syllable quantity. Instead, the model included an additional effect of gender, showing later peaks for the three male participants.

When measured relative to the beginning of the word, there was no reliable difference between the two types of first syllables in this subset. In contrast, relative to this segmental anchor point, peaks were significantly later in narrow focus and trisyllabic words. The best model given in Table J.5 however included a significant interaction between initial syllable type and number of syllables, indicating that the difference between CV.X and CVVC.X words was larger for trisyllabic than for disyllabic words.

Table J.2: Coefficients of the best linear mixed-effects model of the distance of the peak from the beginning of the first syllable nucleus (in ms) for disyllabic words, with random by-item and random by-subject effects of focus and trial.

	Estimate	Std. Error	t value
(Intercept)	36.0330	9.3788	3.8419
1st syllable CVC	6.6054	10.2541	0.6442
1st syllable CVC_g	26.6318	12.5645	2.1196
1st syllable CVV	−4.0275	10.8658	−0.3707
1st syllable CVVC	−6.1588	10.3204	−0.5968
2nd syllable Q2	16.9573	7.4054	2.2899
Narrow focus	17.6225	7.7454	2.2752
Trial	−0.0132	0.0324	−0.4065
List 2	−13.6831	4.9784	−2.7485
List 3	−1.4402	4.9487	−0.2910
1st syll. CVC : 2nd syll. Q2	−7.0157	8.5657	−0.8190
1st syll. CVC_g : 2nd syll. Q2	−34.9874	10.5121	−3.3283
1st syll. CVV : 2nd syll. Q2	−17.7098	9.1102	−1.9440
1st syll. CVVC : 2nd syll. Q2	−10.1369	8.5414	−1.1868
1st syll. CVC : Narrow focus	−5.5809	8.4608	−0.6596
1st syll. CVC_g : Narrow focus	−27.8895	10.3303	−2.6998
1st syll. CVV : Narrow focus	−8.1745	9.0215	−0.9061
1st syll. CVVC : Narrow focus	−9.4512	8.4812	−1.1144
2nd syll. Q2 : Narrow focus	−9.3134	10.4274	−0.8932
1st syll. CVC : 2nd syll. Q2 : N. foc.	1.4282	12.0778	0.1182
1st syll. CVC_g : 2nd syll. Q2 : N. foc.	37.8720	14.6450	2.5860
1st syll. CVV : 2nd syll. Q2 : N. foc.	10.1451	12.7940	0.7930
1st syll. CVVC : 2nd syll. Q2 : N. foc.	−0.4224	12.0459	−0.0351

Table J.3: Coefficients of the best linear mixed-effects model of the distance of the peak from the beginning of the first syllable nucleus (in ms) for contrast set 1, with random by-item effects of second syllable quantity and random by-subject slopes.

	Estimate	Std. Error	t value
(Intercept)	41.2584	5.8188	7.0905
1st syllable CVVC	−16.3603	5.1956	−3.1489
Trisyllabic	8.7516	4.0716	2.1494
Narrow focus	10.4078	3.3258	3.1295
2nd syllable Q2	2.7278	5.1147	0.5333

Table J.4: Coefficients of the best linear mixed-effects model of the distance of the peak from the end of the first mora (in ms) for contrast set 1, with random by-item and by-subject slopes.

	Estimate	Std. Error	t value
(Intercept)	−51.8116	5.0433	−10.2733
1st syllable CVVC	−16.8379	4.9948	−3.3711
Trisyllabic	8.6110	3.4743	2.4785
Male gender	25.3134	10.8010	2.3436

Table J.5: Coefficients of the best linear mixed-effects model of the distance of the peak from the beginning of the word (in ms) for contrast set 1, with random by-item and by-subject slopes.

	Estimate	Std. Error	t value
(Intercept)	121.9260	7.8252	15.5811
1st syllable CVVC	−10.6602	8.2335	−1.2947
Trisyllabic	17.0537	5.4660	3.1200
Narrow focus	15.5237	3.6929	4.2036
1st syllable CVVC : trisyllabic	−15.8264	7.8013	−2.0287

For the subset of data from contrast set 2 (first syllable CVV vs. CVVC, N = 521), models of peak distance from the beginning of the first syllable nucleus and from the end of the first mora did not provide any evidence for significant effects of any experimental factor. None of the compared models exhibited t-values over 2. More importantly, models including the experimental variables provided no better fit than the respective null models specifying only random by-subject and by-item slopes. Only for the least local measure of alignment, the distance of the peak from the beginning of the word, did a model outperform the null model. As shown in Table J.6, it indicated a positive effect of narrow focus, while adding further factors did not improve the fit. This suggests that peaks were later in narrow focus condition also in this subset, whereas differences in number of syllables, second syllable vowel quantity and first syllable structure were not relevant for peak alignment.

Table J.6: Coefficients of the best linear mixed-effects model of the distance of the peak from the beginning of the word (in ms) for contrast set 2, with random by-item and by-subject slopes.

	Estimate	Std. Error	t value
(Intercept)	104.9040	4.6647	22.4887
Narrow focus	12.1425	3.8622	3.1439

In the subset of data from contrast set 3 (first syllable CVC vs. CVVC, N = 519), peak alignment relative to all three segmental landmarks was better accounted for by models containing predictors than by null models. The best model of the distance from the beginning of the first syllable nucleus included effects of the factors number of syllables and trial, indicating that trisyllabic words had later peaks and peaks were realised slightly earlier for items placed later in the experimental session (cf. Table J.7). With regard to the end of the first mora, there was again statistical evidence for later peaks in trisyllabic words and, additionally, in words with Q2 vowels in the second syllable (cf. Table J.8)[1]. The best model of the distance of the peak from the beginning of the word suggested significant positive effects of number of syllables and focus condition and a negative effect of trial (cf. Table J.9). None of the three measures of alignment indicated a significant difference between initial CVC and CVVC syllables.

For the subset of data from contrast set 4 (first syllable CVC vs. CVCg, N = 523), the models of the distance of the peak from the beginning of the

[1]An alternative model without number of syllables as a predictor variable only showed a marginally worse fit. However, in contrast to the model reported in Table J.8, it was not significantly different from a null model according to an ANOVA comparison.

Table J.7: Coefficients of the best linear mixed-effects model of the distance of the peak from the beginning of the first syllable nucleus (in ms) for contrast set 3, with random by-item and by-subject slopes.

	Estimate	Std. Error	t value
(Intercept)	32.7388	5.2220	6.2694
Trisyllabic	7.1949	3.4268	2.0996
Trial	−0.0928	0.0433	−2.1420

Table J.8: Coefficients of the best linear mixed-effects model of the distance of the peak from the end of the first mora (in ms) for contrast set 3, with random by-item and by-subject slopes.

	Estimate	Std. Error	t value
(Intercept)	−77.4203	5.7030	−13.5755
Trisyllabic	14.7360	4.0418	3.6459
2nd syllable Q2	8.1813	4.0870	2.0018

Table J.9: Coefficients of the best linear mixed-effects model of the distance of the peak from the beginning of the word (in ms) for contrast set 3, with random by-item and by-subject slopes.

	Estimate	Std. Error	t value
(Intercept)	105.5781	5.7278	18.4324
Trisyllabic	7.7908	3.6443	2.1378
Narrow focus	9.5474	3.4595	2.7598
Trial	−0.1270	0.0461	−2.7527

first syllable nucleus and from the beginning of the word both suggested later timing in narrow focus as the only significant effect (cf. Tables J.10 and J.11). For the distance of the peak from the end of the first mora, no indication for any significant effects were found, since models specifying the experimental variables as predictors did not show a significantly better fit than a null model.

Table J.10: Coefficients of the best linear mixed-effects model of the distance of the peak from the beginning of the first syllable nucleus (in ms) for contrast set 4, with random by-item and by-subject slopes.

	Estimate	Std. Error	t value
(Intercept)	48.9531	9.6757	5.0594
Narrow focus	8.2208	3.1145	2.6395

Table J.11: Coefficients of the best linear mixed-effects model of the distance of the peak from the beginning of the word (in ms) for contrast set 4, with random by-item and by-subject slopes.

	Estimate	Std. Error	t value
(Intercept)	119.7711	4.708	25.4401
Narrow focus	10.0149	3.268	3.0645

Finally, all models of peak alignment in the subset of data from contrast set 5 (first syllable CVC vs. CVV, $N = 523$) pointed to later peaks in narrow focus words. While this was the only significant effect for the distance of the peak from the beginning of the word (cf. Table J.14), relative to the beginning of the first syllable nucleus peaks were also earlier in words starting with CVV than in those starting with CVC syllables (cf. Table J.12). Relative to the end of the first mora, there was no evidence for an effect of first syllable structure. Instead, the model showed evidence for significantly later peaks in trisyllabic words and targets with long vowels in the second syllable (cf. Table J.13).

Table J.12: Coefficients of the best linear mixed-effects model of the distance of the peak from the beginning of the first syllable nucleus (in ms) for contrast set 5, with random by-item and by-subject slopes.

	Estimate	Std. Error	t value
(Intercept)	43.0885	5.4374	7.9245
1st syllable CVV	−17.7454	5.3056	−3.3447
Narrow focus	12.9893	2.6709	4.8632

Table J.13: Coefficients of the best linear mixed-effects model of the distance of the peak from the end of the first mora (in ms) for contrast set 5, with random by-item and by-subject slopes.

	Estimate	Std. Error	t value
(Intercept)	−76.5741	5.7361	−13.3496
Trisyllabic	9.2087	3.4246	2.6890
Narrow focus	9.4854	2.7226	3.4840
2nd syllable Q2	6.7831	3.3402	2.0307

Table J.14: Coefficients of the best linear mixed-effects model of the distance of the peak from the beginning of the word (in ms) for contrast set 5, with random by-item and by-subject slopes.

	Estimate	Std. Error	t value
(Intercept)	110.9928	7.3736	15.0528
Narrow focus	17.7117	2.8939	6.1204

Appendix K

Background information on participants of the perception experiment

This appendix provides information about the subjects participating in the perception experiment discussed in Chapter 5.

Table K.1: List, age, sex, university study subject(s) and unknown items for participants of the perception experiment. The four lists differed in presentation order of the trials. The last column lists experimental target or filler items that the participant was not familiar with.

Id	List	Age	Sex	Study subject	Unknown word
01	1	23	f	anthropology	none
02	2	25	f	economic and social history	none
03	3	24	f	French philology	none
04	4	24	f	history	none
05	1	28	f	social psychology	none
06	2	20	f	general literature	none
07	3	47	f	French philology	none
08	4	25	m	musicology	kylmäkoje
09	1	32	f	African studies	none
10	2	27	m	theoretical philosophy	none
11	3	25	f	East Asian studies	none
12	4	22	f	folkloristics	none
13	1	23	m	musicology	none
14	2	22	f	development studies	none
15	3	19	f	Finnish literature	none
16	4	18	m	physics	none
17	1	24	m	East Asian studies	kissansilmä, karhunkieli
18	2	34	m	cultural studies	none
19	3	31	f	history	none
21	4	21	f	Russian language and literature	none
22	1	25	f	English philology	none
23	2	23	f	French philology	koiranleuka
24	3	19	f	Spanish philology	puuviilu (filler)
25	4	20	m	practical philosophy	none
26	1	20	f	French philology	puuviilu (filler), (karhunkieli)
27	2	25	f	English philology	none
28	3	32	f	English philology	none

Table K.2: Dialect and places of residency of participants of the perception experiment.

Id	Dialect	Lived in
01	standard	Vantaa, Nurmijärvi, Helsinki
02	standard, maybe a sort of Helsinki accent	Helsinki
03	standard and Helsinki dialect	France, Helsinki
04	standard	Helsinki, Budapest
05	standard	Helsinki, Järvenpää
06	standard	Lohja, Helsinki
07	standard	Helsinki, Veikkola, Kirkkonummi, France
08	standard	Helsinki, Vantaa
09	standard	Helsinki, Porvoo, Brussels
10	standard	Kerava, Helsinki
11	standard	Helsinki, Kauniainen, Espoo
12	standard	Helsinki, Espoo
13	standard	Espoo, London, Helsinki
14	standard and / or Helsinki dialect	Helsinki, Uganda
15	standard	Tampere, London, Helsinki, Vantaa
16	standard	Helsinki, Jokela
17	standard	Uppsala, Helsinki, Vantaa, Shanghai
18	standard	Helsinki, St. Petersburg
19	standard	[unreadble]
21	standard	Helsinki
22	standard	Vantaa, Helsinki, Cardiff
23	standard	Espoo, Helsinki
24	standard	Nurmijärvi, Helsinki
25	standard or youth slang (nykyslangia)	Helsinki
26	standard	Espoo
27	standard	Helsinki, Loviisa, Nottingham
28	standard	Helsinki, Hämeenlinna, Bern

Table K.3: Native language(s) and foreign languages spoken by participants of the perception experiment.

Id	Native language	Other languages spoken
01	Finnish	Swedish, English, German, (Spanish, Portuguese)
02	Finnish	Swedish, English, German, Spanish
03	Finnish, father's language French	English, Swedish, Spanish, Russian, German
04	Finnish	English, Swedish, French, (Hungarian, Latin, German)
05	Finnish	Swedish, English, Spanish
06	Finnish	Swedish, English, (French, Spanish)
07	Finnish	French, English, Swedish, Spanish, Portuguese, Russian
08	Finnish	English, Swedish, German
09	Finnish	French, English, Swedish
10	Finnish	English, Swedish, French, Latin, Greek
11	Finnish	English, Japanese, (Swedish)
12	Finnish	English, Spanish, Swedish, French, Hungarian
13	Finnish	English, French, Swedish
14	Finnish	English, French, Swedish
15	Finnish	English, Swedish, French
16	Finnish	English, Swedish
17	Finnish	English, French, Swedish, German, Mandarin, Korean, Cantonese
18	Finnish	English, Swedish, French, Spanish, German, Russian, Latin
19	Finnish	English, Swedish, German
21	Finnish	Russian, Ukrainian, English, Swedish, German
22	Finnish	English, Swedish, French, Spanish, (Japanese)
23	Finnish	Swedish, English, French, Italian, German
24	Finnish	Spanish, English, Swedish
25	Finnish	English, Swedish
26	Finnish	English, French, Swedish, German, Spanish, Chinese
27	Finnish	English, Swedish, Spanish
28	Finnish	English, Swedish, French, Italian, German, Spanish, (Japanese)

Appendix L

List of target and filler items of the perception experiment

This appendix lists the target and filler items used in the study discussed in Chapter 5. All lexical items used in the perception experiment—excluding the practice trials—are shown in the table on the next four pages. Filler items are displayed in the same line as the target item that they were matched to by compound lemma frequency per million (cp. frequency).

Table L.1: List of target and filler items for perception experiment.

Target item	Translation (compound)	Translation (noun phrase)	Filler compound	Translation	Filler noun phrase	Translation	Target cp. frequency	Filler cp. frequency
amerikanrauta	old-fashioned American car	America's iron	leipälapio	peel (baker's shovel)	lapion varsi	spade handle	0.4928	0.4928
annansilmä	a type of begonia	Anna's eye	mätitahna	roe paste	mätä kasvi	rotten vegetable	0.0870	0.0870
anopinkieli	a plant (Sansevieria)	mother-in-law's tongue	lasikenkä	glass shoe	Lassen kenkä	Lasse's shoe	0.0580	0.0580
hapankaali	sauerkraut	sour / rotten cabbage	heinäsirkka	grasshopper	Sirkan kenkä	Sirkka's shoe	1.3623	1.3623
harakanvarvas	scrawl	magpie's toe	kuivahiiva	dry yeast	kuiva pulla	dry cardamom bread	0.2029	0.2029
hiidenkirnu	pothole	demon's churn	parsakaali	broccoli	parsan varsi	stalk of asparagus	0.6377	0.6377
hiirenkorva	a newly unfolded leaf of a tree	mouse's ear	mustekala	squid	musta kala	black fish	0.6667	0.6667
hullunmylly	bedlam (ridiculously chaotic situation)	crazy person's mill	pehmolelu	cuddly toy	pehmeä tyyny	soft pillow	0.9855	0.9855

Continued on next page

Table L.1 – *Continued from previous page*

Target item	Translation (compound)	Translation (noun phrase)	Filler compound	Translation	Filler noun phrase	Translation	Target cp. frequency	Filler cp. frequency
isokenkä	person with a lot of power	big shoe	puulasta	spatula (from wood)	puun lusto	growth ring of a tree	0.0870	0.0870
isovarvas	the big toe	a big toe	hiilihanko	poker	uunin luukku	oven's hatch	0.2899	0.2899
Joensuu	city in North Karelia	estuary	kesämökki	summer cottage	punainen vaja	red hut	45.6828	47.1806
juoponnappi	clothes buttoned up the wrong way	drunkard's button	saunaharja	sauna brush	saunan kiuas	oven of the sauna	0.0290	0.0290
karhunkieli	scouring pad	bear's tongue	pappamopo	oldfashioned type of moped	vaarin pyörä	grandfather's bicycle	0.0870	0.0870
ketunleipä	plant (wood sorrel)	fox's bread	saamelaisasu	Saami traditional costume	saamelainen kota	Saami tipi-like hut	0.0290	0.0290
kissankello	flower (harebell)	cat's bell	metsäjänis	mountain hare	ruskea kani	brown rabbit	0.6667	0.6667
kissansilmä	rear reflector on a bike	cat's eye	lumiluola	snow cave	lumen luoja	sombody shovelling snow	0.0870	0.0870
koiranleuka	practical joker	dog's chin	käsinukke	hand puppet	Katin nukke	Kati's doll	0.3768	0.3768
koiranputki	plant (cow parsley)	dog's pipe	nallipyssy	toy pistol	pyssyn kuula	pistol's bullet	0.5217	0.5217

Continued on next page

Table L.1 – *Continued from previous page*

Target item	Translation (compound)	Translation (noun phrase)	Filler compound	Translation	Filler noun phrase	Translation	Target cp. frequency	Filler cp. frequency
kurjenpolvi	plant (geranium)	crane's knee	autonakku	car battery	auton moottori	motor of a car	0.1159	0.1159
kylmäkoje	fridge / freezer	cool device	kuusenjalka	Christmas tree stand	kuusen kanto	stump of a spruce tree	0.0290	0.0290
märkäpuku	diving suit	wet suit	sivupeili	wing mirror	peilin kehys	mirror frame	0.3478	0.3478
mustalammas	black sheep (a person)	black sheep	teräsvilla	steel wool	terän viilaus	filing a blade	0.0290	0.0290
outolintu	extravagant person	strange bird	tomusokeri	icing sugar	sokerin pala	piece of sugar	0.4928	0.4928
pikkutakki	blazer jacket	small coat	raitiovaunu	tram car	vihreä juna	green train	1.4493	1.4493
pitkäkynsi	thief	long (finger or toe) nail	puuviilu	veneer	puinen viulu	wooden violin	0.0580	0.0580
pojannaskali	cheeky monkey	the boy's awl	rautaovi	iron door	rautainen lukko	iron lock	0.0580	0.0580
poninhäntä	ponytail	pony's tail	autiotupa	unlocked wilderness cabin free to use	autio saari	uninhabited island	0.9275	0.9275
pystykiekko	icehockey pass along the long side of the field	erect puck	vesikannu	ewer	veden kanto	carrying water	0.0290	0.0290

Continued on next page

Table L.1 – *Continued from previous page*

Target item	Translation (compound)	Translation (noun phrase)	Filler compound	Translation	Filler noun phrase	Translation	Target cp. frequency	Filler cp. frequency
pystykorva	spitz	erect ear	käyrätorvi	french horn	käyrä torni	crooked tower	4.7246	4.2609
sudenpentu	junior scout	wolf cub	pannukakku	pancake (made in the oven)	pannun kahva	handle of a pot / pan	1.0725	0.9855
suorakaide	rectangle	straight balustrade	koivuhalko	birch log	puun oksa	branch of a tree	1.4493	1.3623
vanhapiika	old maid (unmarried woman)	old maid-servant or female farm hand	kaasugrilli	gas grill	kaareva rillit	curved glasses	0.4928	0.4928
verenpisara	flower (fuchsia)	drop of blood	kääntösilta	pivot bridge	kivinen silta	stone bridge	0.2319	0.2319
villisika	wild boar	wild pig	paloauto	fire engine	Panun auto	Panu's car	6.2319	5.5072

Appendix M

Abstract

This dissertation provides an analysis of Finnish prosody, with a focus on the sentence or phrase level. The thesis analyses Finnish as a phrase language. Thus, it accounts for prosodic variation through prosodic phrasing and explains intonational differences in terms of phrase tones.

Finnish intonation has traditionally been described in terms of accents associated with stressed syllables, i.e. similarly as prototypical intonation languages like English or German. However, accents are usually described as uniform instead of forming an inventory of contrasting accent types. The present thesis confirms the uniformity of Finnish tonal contours and explains it as based on realisations of tones associated with prosodic phrases instead of accents. Two levels of phrasing are discussed: Prosodic phrases (p-phrases) and intonational phrases (i-phrases). Most prominently, the p-phrase is marked by a high tone associated with its beginning and a low tone associated with its end; realisations of these tones form the rise-fall contours traditionally analysed as accents. The i-phrase is associated with a final tone that is either low or high and additionally marked by voice quality and final lengthening. While the tonal specifications of these phrases are thus predominantly invariant, variation arises from different distributions of phrases.

This analysis is based on three studies, two production experiments and one perception study. The first production study investigated systematic variation in information structure, first syllable vowel quantity and the target word's position in the sentence, while the second production experiment induced variation in information structure, first and second syllable type and number of syllables. In addition to fundamental frequency, the materials were analysed regarding duration, the occurrence of pauses and voice quality. The perception study investigated the interpretation of compound/noun phrase minimal pairs with manipulated fundamental frequency contours using a two-alternative forced-choice picture selection task. Additionally, a

pilot perception study on variation in peak height and timing supported the assumption of uniform tonal contours.

Appendix N

Deutsche Zusammenfassung

Diese Dissertation bietet eine Analyse der Prosodie des Finnischen auf der Satzebene. Die Dissertation beschreibt Finnisch als eine Phrasensprache, d.h. eine Sprache, deren prosodische Variation durch Phrasierung und mit prosodischen Phrasen assoziierte Phrasentöne bestimmt wird.

Bisherige Untersuchungen der finnischen Intonation haben diese als durch Akzente, die mit betonten Silben assoziiert sind, beschrieben, d.h. ähnlich wie typische Intonationssprachen wie Englisch und Deutsch. Diese Akzente werden jedoch übereinstimmend als uniform, also vom gleichen Typ, dargestellt, so dass kein Inventar an kontrastierenden Akzenten aufgestellt werden konnte. Die vorliegende Dissertation bestätigt die tonale Uniformität von Kontouren des Finnischen, die als Realisierungen von Phrasentönen beschrieben werden. Dabei nimmt die Arbeit auf zwei Phrasierungsebenen Bezug: prosodische Phrasen (p-Phrasen) und Intonationsphrasen (i-Phrasen). P-Phrasen werden hauptsächlich durch zwei Töne markiert, einen hohen Ton am Anfang der Phrase und einen tiefen Ton am Ende. Realisierungen dieser Töne sind für die steigend-fallenden Konturen verantwortlich, die traditionell als (uniforme) Akzente beschrieben wurden. I-Phrasen enden mit einem hohen oder tiefen Ton und weisen weitere Markierungen wie finale Längung und nicht-modale Stimmqualität auf. Während die tonale Markierung von prosodischen Phrasen weitgehend gleichbleibend ist, entsteht prosodische Variation durch Unterschiede in der Distribution der prosodischen Phrasen.

Die hier vorgestellte Analyse beruht im Wesentlichen auf drei Studien, zwei Produktionsexperimenten und einer Perzeptionsstudie. Die erste Produktionsstudie untersuchte systematische informationsstrukturelle Variation, sowie Variation in der Quantität des Vokals der ersten Silbe und der Position des betrachteten Wortes im Satz. Das zweite Experiment erzielte Variation in der Struktur der ersten und zweiten Silbe, der Silbenanzahl und Informationsstruktur des Satzes. Zusätzlich zur Grundfrequenz wurden Dauer, Pausenvorkommen und Stimmqualität untersucht. Die Perzeptionsstudie erkundete die Interpretationen von Minimalpaaren mit manipulierten Grundfrequenzkonturen, die entweder als Komposita oder als Nominalphrasen verstanden werden konnten. Eine zusätzliche Pilotstudie zur Wahrnehmung von Höhe und Timing von tonalen Gipfeln lieferte weitere Unterstützung für die Annahme tonaler Uniformität.

Appendix O

Lebenslauf

Anja-Helene Arnhold

Studium und Ausbildung

10/2010–06/2013	Johann Wolfgang Goethe-Universität Frankfurt am Main Promotionsstudium Kognitive Linguistik Betreuerin: Prof. Dr. Caroline Féry
09/2009–07/2011	Universität Helsinki, Finnland Phonetik, Gaststudentin
10/2007–09/2010	Universität Potsdam Promotionsstudium Allgemeine Sprachwissenschaft, Betreuerin: Prof. Dr. Caroline Féry
10/2002–10/2007	Universität Potsdam Allgemeine und Theoretische Linguistik, Literaturwissenschaft (Germanistik) Abschluss: Magister Artium
08/2005–06/2006	Roskilde Universitetscenter, Dänemark Kulturwissenschaft und Germanistik, Austauschstudentin
06/2002	Abitur, Beethoven-Schule Berlin (Gymnasium)

Berufserfahrung

seit 09/2013	Universität Alberta, Institut für Linguistik Postdoctoral fellow und Vertragsdozentin
03/2010–05/2010	Universität Helsinki, Fakultät für Verhaltens-wissenschaften Vertragsdozentin
09/2011–10/2011	Universität Helsinki, Fakultät für Verhaltens-wissenschaften Vertragsdozentin
10/2007–03/2009	Universität Potsdam, Institut für Linguistik, DFG-Projekt FA 255/5 'Morphosyntax und Phonologie von diskontinuierlichen Nominal- und Präpositionalphrasen' Projektleiter: Prof. Dr. Gisbert Fanselow und Prof. Dr. Caroline Féry Wissenschaftliche Mitarbeiterin
08/2006–09/2007	Universität Potsdam, Institut für Linguistik Studentische Hilfskraft
08/2003–07/2005	Universität Potsdam, Institut für Linguistik, Sonderforschungsbereich '632: Informationsstruktur', Projekt 'D2: Typologie der Informationsstruktur' Projektleiter: Prof. Dr. Caroline Féry, Prof. Dr. Gisbert Fanselow und Prof. Dr. Manfred Krifka Studentische Hilfskraft