On the pragmatic and semantic functions of Estonian sentence prosody

Nele Salveste

München 2015

On the pragmatic and semantic functions of Estonian sentence prosody

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophie
an der Ludwig-Maximilians-Universität München

vorgelegt von
Nele Salveste
aus Järva-Madise, Estland
2015

Erstgutachter: Prof. Dr. Jonathan Harrington

Zweitgutachter: Prof. Dr. Phil Hoole

Tag der mündlichen Prüfung: 7. Juli. 2015

Acknowledgments

I'm deeply grateful to my supervisor, Jonathan Harrington, who has guided and supported me throughout my doctoral studies. Through his questions and comments on my work he has demonstrated to me a clear, logical and rigorous way of scientific thinking. While I may not yet have perfected such thinking in this doctoral work, I aim to pursue it in my future work.

Another person to whom I am immensely grateful is Felicitas Kleber. Without her help and advice on experiments, on writing and on everyday academic life I would not have come so far.

I'm extremely thankful to, again, Jonathan Harrington and Felicitas Kleber, but also to Sanjida Sattar-Loveday, Eva Liina Asu-Garcia, Jessica Siddins, Hanna Ruch, Piia Taremaa and Miina Norvik for reading and correcting my rough drafts and making sense out of my confusingly presented ideas. In addition, I would like to express my gratitude to Clara Tillmanns for helping to correct and write the German summary for the thesis.

I appreciate Pire Teras' and Jana Häussler's help reading the sketches of experiments and giving me extremely valuable advice. I'm very thankful to Pärtel Lippus, Klaus Jänsch and Christoph Draxler who have helped and assisted me in running the experiments. I thank all the participants of my experiments, as I would not have had anything to report without them. The analyses of acoustic data have gained a lot from the frequent discussions of my data with Frank Kügler and Susanne Genzel during my stay at the University of Potsdam. Their interest and enthusiasm towards my studies motivated me a lot.

Finally, the thesis gained a lot from the reviews by Jonathan Harrington and Phil Hoole. I'm also very grateful for the valuable comments from Elena Skribnik.

The dissertation would not have been possible without the scholarships I have been awarded. My research was mainly supported by the national scholarship program Kristjan Jaak, which is funded and managed by Archimedes Foundation in collaboration with the Estonian Ministry of Education and Research. I received additional support from the Estonian Science Foundation grant no. 7904 and the Estonian Research Council grant IUT2-37. Finally, I was very happy to be a scholarship holder of a short-

term grant by the Collaborative Research Centre SFB 632 at the University of Potsdam during the academic year of 2013/14.

I owe my great gratitude to my friendly and supportive colleagues and friends at the Institute of Phonetics and Speech Processing of the University of LMU Munich and at the Phonetics Lab of the University of Tartu.

My very special gratitude for love and support belongs to my sister – Liina. She was always willing to share her clear and concise intuitions about our mother tongue and to take part in my experiments. I thank my best and dearest friend Rauno for his understanding, helping and loving company.

My doctoral studies have taken me on a wonderful (time to time also a difficult) journey and paved the way for a number of further studies.

Contents

List of figures

Abstract

The goal of the dissertation was to investigate intonational correlates of *information structure* in a *free word order* language, Estonian. Information-structural categories such as *focus* or *givenness* are expressed by different grammatical means (e.g. pronoun, presence of accent, word order etc.) in different languages of the world (Chafe, 1976; 1987; Prince, 1981; 1992; Lambrecht, 1994; Gundel, 1999). The main cue of focus in intonation languages (e.g. English and German) is pitch accent (Halliday, 1967a; Ladd, 2008). In free word order languages, information structure affects the position of words in a sentence (É. Kiss, 1995) and sometimes it is even implied that word order in a free word order language might function like pitch accent in an intonation language (Lambrecht 1994: 240). The study reports on perception and production experiments on the effects of focus and givenness on Estonian sentence intonation. The aim of the experiments was to establish whether information structure has tonal correlates in Estonian, and if so, whether information structure or word order interacts more strongly with sentence intonation.

A perception experiment showed that L1-Estonian listeners perceive pitch prominence as focus and accent shift as a change of sentence focus. A speech production study showed congruently that L1-Estonian speakers do use accent shift, and mark sentence focus with pitch accent. Another speech production experiment demonstrated that there is no phonetic difference between new information focus (e.g. "*What did Lena draw?*" – "*Lena drew a whale.*") and corrective focus (e.g. "*Lena drew a lion.*" – "*No! She drew a whale*"). The last experiment showed that given information is signalled with varying F0 range, if followed by focus, but without a pitch accent, if preceded by focus.

All the experiments revealed that word order has a weak influence on sentence intonation. Sentence intonation interacts with focus and givenness in Estonian. As a conclusion, it is suggested that the pragmatic functions of word order, which apparently can be overridden by focus interpretation, are slightly different from the functions of pitch accent.

1. Introduction

Information flow in speech is controlled by information structure. Comprehension of new relevant information is considerably easier if some background information is already known. Consider, for instance, what might cause disruption in understanding the example in (1.1).

(1.1)

> I know Tiia. She is a good friend of mine. Tiia teaches a language course. It is her mother tongue and she enjoys teaching it.

The stream of information in this example should go from the person named Tiia to a language course she teaches and then to her mother tongue. However, some of the new information provided in the last sentence (mother tongue) is related to information that has not yet been established. Besides the language course, no specific language has been mentioned up to that point and this may cause confusion. A successful transmission of information therefore includes not just new information but also *given/old information* to which *new information* is attached. Old information and new information form a part of the so-called *information structure* that is necessary to parse the information flow. The information structure of a sentence within a discourse can be recognized based on certain linguistic features. Observe in (1.1) that full nouns (*Tiia*, *friend of mine*, *language course*, *mother tongue*) introduce new information, whereas the pronoun *it* refers to the information already established in the discourse.

The thesis is organized as follows. First chapter discusses the information structure and the fine structure of old and new information as it appears expressed in the languages of the world. This fine structure causes complex relation between linguistic categories such as *focus* (Halliday, 1967a; Lambrecht, 1994; Gundel, 1999) or *givenness* (Chafe, 1976; 1987; Lambrecht, 1994; Gundel, 1999; Baumann, 2006). The second and the third subsection of the first chapter discuss two formal means of the information structure: sentence accent and word order. The fourth subsection introduces the object of the study: the Estonian language. Chapters 2, 3, 4 and 5 report on the experiments carried out on the intonational encoding of information structure in

Estonian. Chapter 6 provides a tentative interpretation of the experimental results in the framework of the chosen theory of information structure.

1.1. Information structure

Information structure has gained a lot of attention in linguistic research (see Vallduví (1993) for a comprehensive overview). Numerous definitions in literature (see Baumann (2006: 41–42) for a complete list of examples) try to capture the distinction between old and new information and linguistic means by which they are expressed in a language. There is a tradition to speak about *theme-rheme* structure (Halliday, 1967a; Firbas, 1966; Daneš, 1966) about *topic-comment* structure (Reinhart, 1982; Gundel, 1985) and about *topic-focus* structure (Sgall et al., 1986, Lambrecht, 1994). Another branch of information-structure research concentrates on terms *given vs. new* (Chafe, 1976; Brown, 1983; Terken & Hirschberg, 1984; Xu and Xu, 2005; Cruttenden, 2006; Baumann, 2006; Baumann and Grice, 2006; Baumann and Riester, 2012), just on *focus* (Gussenhoven, 2007; Ladd, 2008) or on *ground-focus* articulation (Vallduví and Engdahl, 1996). Even if two accounts use the same terminology, they often differ slightly in the exact definitions of the concepts depending on whether the theoretical framework relies on syntactic, pragmatic or phonological investigation of information structure. The syntactic approaches often operate with the terminological pairs of topic-comment or topic-focus, whereas the phonology research prefers focus and givenness. The definitions of givenness and focus in phonology research can be traced back to theories of information structure developed in Halliday (1967ab) and in Chafe (1976) but recent studies on sentence intonation (e.g. Breen et al., 2010) assume alternative semantics account of focus proposed by Rooth (1992).

The theoretical framework of the present dissertation is developed based on the definitions of focus and givenness (Halliday, 1967b; Rooth, 1992; Lambrecht, 1994; Gundel, 1999; Chafe, 1976). The main idea of the thesis is that information structure can be observed on two superimposed structures. The first structure is called theme-rheme structure (Halliday, 1967a). The second structure is a discourse-related structure that affects the informational status of sentence constituents in terms of givenness (Chafe, 1976; Lambrecht, 1994) and focus (Halliday, 1967a; Rooth, 1992; Lambrecht, 1994; Gundel, 1999). It is to be noted that the use of the term focus in this thesis differs essentially from the use of focus paired with topic (Sgall et al., 1986). In some studies,

topic-focus structure is assumed to interact with *contrast* (see e.g. Repp, 2010) in a similar way as focus is defined to interact with theme-rheme structure or with activeness and identifiability in the current study. Contrast is treated synonymously with correction and is defined as a subtype of focus (Krifka, 2007; Gussenhoven, 2007) in chapter 4.

1.1.1. Focus

A theory by Michael A. K. Halliday (1967ab) distinguishes between *given* and *new*, whereas a theory by Knud Lambrecht (1994) prefers to differentiate between *topic*, *focus, activeness* and *identifiability*.

According to Halliday (1967a), the given-new dichotomy is a discourse structure "through which the speaker both organizes the act of communication into a chain of message blocks, the 'information units', and specifies within each message block the value of the components in the progression of the discourse" (1967a: 211). Halliday (1967ab) defines new information on the basis of intonation: "Information structure is realized phonologically by 'tonality', the distribution of the text into tone groups: one information unit is realized as one tone group" (Halliday 1967a: 200). Notably, "the information unit is what the speaker chooses to encode as a unit of discourse" (Halliday 1967a: 202). For each information unit that is a tone group, there is one "point of information focus", maximally two (Halliday, 1967a: 204), that correspond to "one obligatory component, the 'tonic segment', and one optional component, the 'pretonic segment'" (Halliday 1976a: 203). In principle, new information in Halliday's (1967ab) account is a prosodic prominence that "is a matter of pitch (pitch movement, not pitch level) and secondarily one of duration and intensity" (Halliday 1967a: 203). Thus, prosodic prominence or pitch accent correlates with new information and this aspect of Halliday's approach has received much of the criticism in later literature (e.g. Lambrecht, 1994; Gundel, 1999).

In Halliday's definition, focus is a part of sentence that the speaker presents as new to the hearer: "The focus of the message, it is suggested, is that which is represented by the speaker as being new, textually (and situationally) non-derivable information" (Halliday 1967a: 205). Lambrecht (1994) adopts Halliday's definition (1967a) of *focus*: "the focus is that portion of a proposition which cannot be taken for granted at the time of speech" (Lambrecht, 1994: 207). For Lambrecht it is important to recognize that focus is a pragmatic relation between entities expressed in an utterance.

He describes sentences uttered in a context in which it is impossible to point to a particular sentence constituent that is in focus. Instead, the relation between the entities is in focus. See example (1.2) for illustration.

(1.2)

I did it, because you're my friend! (Lambrecht 1994: 58)

In (1.2) the knowledge that the speech participants share is the fact that the person referred to as *I* did something and the fact that this person and the addressee are friends but the relation (*because*) between these two facts is unexpected and 'new information' for the hearer. Thus, according to Lambrecht, focus is a pragmatic relation that is established within a contextually bound utterance between the referring expressions involved in an utterance (1994: 210). To put it more simply, focus is a pragmatic role that the referent can play in a text-internal world of discourse (Lambrecht, 1994: 76).

Focus is not just new information in a discourse (Lambrecht 1994: 209f). It is to be noted that already Halliday defines focus as a part of an utterance that the speaker decides to present as new information (1967a: 205). A consequence of this definition is that focus can occasionally contain information that is already known to the speech participants – old information. Consider example (1.3).

(1.3)

a. John's mother voted for Bill.

b. No, she voted for JOHN. (Schwarzchild, 1999)

In the response of (1.3b), *John* is the focus of the sentence but he is previously (John's mother) mentioned and, therefore, also old or known to the listener.

Lambrecht (1994) diverges from his predecessor Halliday (1967a) in a crucial aspect. He separates the definition of focus from one of its formal means: "accent placement and focus marking are not to be equated" (Lambrecht, 1994: 208). Lambrecht (1994: 214) advices to avoid the term 'focused' "because it tends to blur the distinction between a pragmatic category (focus) and a prosodic category (pitch prominence)." Other formal means of focus can be word order or morphological markers (Kuno, 1972; Kiss, 1995; Lambrecht, 1994; Féry & Krifka, 2008). Lambrecht (1994: 240) finds that in respect to the information structure, the variation in accent placement (English) is equivalent to the variation in word order in free word order

languages (in his example, German). This study diverges from Lambrecht's model in this respect. While word order can be indeed used as a formal means of focus, a number of word order variations is caused by the need to signal old information (e.g. Clark & Haviland, 1977).

Matts Rooth (1992) gives a valuable definition for focus in the framework of alternative semantics (Stechov, 1989). In Rooth's definition, a referent in focus has in addition to its lexical and syntactical meaning the so-called *focus semantic value* that arises through a set of alternatives that it evokes in a sentence. A referent in focus informs about other referents that might alternatively be substituted for the referent. Scalar implicature and a question-answer pair demonstrate the essentials of the analysis that results from Rooth's definition. Consider the so-called *partially ordered set* of two propositions about the result of an exam (1.4).

(1.4)

Matts aced > Matts passed (Rooth, 1992)

In (1.4) the lower member (*Matts passed*) of the set is contained in the higher member (*Matts aced*) of the set. Therefore, saying that *Matts passed* implies that his results were not excellent (the mechanism of the so-called scalar implicature). The effect is that an expression in focus "provides information about the underlying [partially ordered] set" (Rooth, 1992: 83) and informs the listener/reader about the alternative (that *Matts aced*) that did not happen.

In respect to the question-answer pairs, "a question determines a set of potential answers" and the focus in an answer signals potential answers (Rooth, 1992: 84), consider example (1.5).

(1.5)

Q: Who cut Bill down to size?

a) Mary cut Bill down to size.

b) Monique cut Bill down to size.

c) Mary cut Björn down to size.

Observe that (a) and (b) are proper answers to the question (Q) and inform about the potential agents for the activity, whereas answer in (c) is incongruent with the question – it does not belong to the set of potential answers that the question determined.

Manfred Krifka (2007: 18) provides a lucid rephrase of Rooth's (1992) formal definition: "Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions." Krifka (2007) agrees with Rooth (1992) but strongly rejects the earlier focus definitions (e.g. Halliday, 1967; Chafe, 1976; Lambrecht, 1994). He considers explications such as 'highlighting', 'most important' and 'new' not quantifiable and, therefore, not satisfactory for a definition. The definition advocated by Krifka (2007) does not seem to outdo the earlier definitions so clearly. Alternatives can be formally determined (in the sense of Rooth, 1992 and Krifka, 2007) as soon as the part of the sentence that a speaker presented as focus (in the sense of Halliday, 1967b and Lambrecht, 1994) is detected. Therefore, in terms of investigating which linguistic means are used for marking focus, the nuances of definition are less relevant. It is more relevant to find a definition of focus that is compatible with different linguistic markings (e.g pitch accent, word order, particles etc.) of focus (Krifka, 2007 as well as Lambrecht, 1994).

On the basis of alternative semantics of focus (Rooth, 1992) and *common ground management* (see Merin, 1994; Groenendijk, 1999; Clark, 1996), Krifka (2007) has developed a fine-grained typology of subtypes of focus, which can well be accommodated with the general idea that focus indicates alternatives of a referent in focus. For example, Krifka differentiates between expression focus and denotational focus, or between semantic and pragmatic uses of focus, or lists different types of focus on the basis of the kind of alternatives focus indicates. Many other focus types he discusses are also pragmatic in their nature: broad focus, narrow focus, verum focus, multiple focus, closed *vs.* open focus (traditionally correction or contrastive *vs.* non-contrastive focus), exhaustive and scalar focus (see Krifka, 2007).

If Krifka develops his theory on the basis of meaning and the types of alternatives, then Gussenhoven (2007) represents a theory of focus types that is mainly derived from the meanings of intonational tunes. Gussenhoven acconts for broad focus and narrow focus, and for presentational, corrective, counterpresuppositional, definitional, contingency and reactivating focus (see Gussenhoven, 2007 for definitions). Gussenhoven's list contains also an identificational focus, but it appears to be expressed by word order rather than by intonation (Gussenhoven, 2007; É. Kiss, 1998). Similarly to focus types offered by Krifka (2007), focus types in Gussenhoven

(2007) are compatible with the general definitions of focus (Halliday, 1967a; Lambrecht, 1994; Rooth, 1992).

From cognitive perspective, Jeanette K. Gundel (1999) argues for three kinds of focus: *psychological focus*, *semantic focus* and *contrastive focus*. Psychological focus is connected to speech participants' attention state. If the focus of attention of all speech participants is on a particular entity, then this entity is assumed to be salient in discourse and, therefore, in (psychological) focus. This definition differs substantially from Halliday's (1967a) or Lambrecht's (1994) accounts of focus. The definition involves entities that are activated (old) in a discourse and, as such, the psychological focus is similar to Chafe's (1976) concept of *givenness* or Lambrecht's concept of *activeness* which will be discussed in the next subsection. The linguistic means for psychological focus according to Gundel (1999: 294) are unstressed personal pronouns, zero anaphors and weakly stressed constituents in general.

Semantic focus is that part of the sentence which relates to the relevant *wh*-question (Gundel, 1999). As such, semantic focus is a relational concept that arises within a structure of a sentence. Every sentence makes a statement about a referent. In this respect, semantic focus resembles Halliday's *theme-rheme* structure. In Halliday's model (1967a), given and new are the informational values that the sentence components can take as the discourse progresses, whereas the theme-rheme structure refers to the sentence structure that is independent from the information flow or discourse context. The *theme-rheme* distinction takes clause as its "point of origin" and structures the clause independently from context (Halliday, 1967a: 212). A clause has a structure of its own right: statement (rheme) and a referent (theme) about what the statement is. The sentence-structural value of components is different from the value they get within the discourse such as being either new or given (Halliday, 1967a).

In her definition, Gundel (1999) differs from Halliday (1967a) by giving the information-structural value also to semantic focus: an entity in semantic focus is new to the discourse. In this sense, semantic focus is parallel with comment or focus in a topic-comment or topic-focus distinction (e.g. Sgall et al., 1986; Reinhart, 1982). However, semantic focus is not sufficient to draw speech participants' attention to an entity (Gundel, 1999: 300). As such, the semantic focus is still rather connected to the structure of a sentence or a clause. According to Gundel (1999), semantic focus is also known to influence the speaker's choice of various linguistic means such as pitch

accent, word order, focus marking particles and different kinds of syntactic structures (e.g. passivization, it-clause).

The contrastive focus in Gundel (1999) corresponds to the focus that Halliday (1967a) and Lambrecht (1994) formalize in their models. For Gundel (1999), an entity is in contrastive focus, if the speaker has drawn listeners' attention to it by some prosodic, syntactic and morphological means. In Gundel (1999), the means of semantic and contrastive focus overlap strongly. The contrastive focus in Gundel (1999) contains entities that can be either new or old to the discourse, similarly to Halliday (1967a) and Lambrecht (1994).

Also, a number of other theoretical accounts refer to contrastivity in the definition of focus (see e.g. Repp, 2010). Rooth (1992) states, for example, that an expression in focus is contrasted to the alternatives that could occupy the same position in an utterance. Being contrastive is therefore an inherent property of focus. However, he aims to "strip away any reference to contrast" in his formal definition (Rooth 1992: 82) and the definition of focus adopted in thesis does not refer to contrast.

1.1.2. Givenness

Chafe (1976: 30) introduces the concept of givenness and accounts for it in terms of consciousness: "Given (or old) information is that knowledge which the speaker assumes to be in the consciousness of the addressee at the time of the utterance." Consequently, the definition implies that the speaker needs and is able to consider the information that the listener already possesses. In Chafe (1976, 1987), the distinction between given and new is like in Halliday (1967ab) – connected mainly to an intonation unit. Chafe (1976: 31) claims that "given information is conveyed in a weaker and more attenuated manner than new information". Interestingly, new information "is not always pronounced with high pitch and strong stress" (Chafe, 1976)

Lambrecht (1994) resumes *givenness* (Chafe 1976: 30) under the concept of *activeness*. Activeness is a property that a referent might have in the minds of the speech participants (Lambrecht, 1994: 76). The concept of activeness corresponds to Chafe's (1974, 1976, 1987) usage of the term *active*. A referent can become activated through mentioning it but it might become activated also through textual or situational inference. Chafe (1987) proposes three *activation states* for the concepts in discourse:

active, *semi-active* and *inactive*. An active concept means that it is in the consciousness of the interlocutors at the moment of a speech event and might be also verbally expressed; a semi-active concept is a concept that can be inferred from either text-external or text-internal context and an inactive concept is neither mentioned in nor inferrable from the context. For the semi-active concepts, Lambrecht (1994: 100) accounts for *textually* accessible, *inferentially* accessible and *situationally* accessible referents. The referents can become textually accessible through a semantic schema or a cognitive frame (Lambrecht, 1994: 99) or by the lexical relations such as hyponyms/hyperonyms, antonyms and synonyms as Stefan Baumann and Martine Grice (2006) have shown.

Lambrecht (1994) finds that activeness, if defined as a linguistic category, cannot successfully account for all the different states of consciousness, which is a psychological phenomenon that is continuous and has an infinite number of states. Activeness as a linguistic category can have only two states: either activated or not. The binary category of activeness "accounts for the relationship between the assumed cognitive states of discourse referents and types of grammatical forms" (Lambrecht 1994: 101). The formal effects of this binary category correspond to the linguistic means of psychological focus in Gundel (1999): a full noun tends to refer to the inactive concept, whereas the pronoun refers to the active referent (cf. example (1.1) above): the activated expressions are encoded without the sentence accent, whereas the inactive expressions with the sentence accent (1994: 107).

Lambrecht (1994) defines activeness as a binary linguistic category, whereas Chafe (1976, 1987) sees it as a continuum. Stefan Baumann and his colleagues (Baumann, 2006; Baumann and Grice, 2006; Baumann and Riester, 2012) have developed the idea of different degrees of givenness further by investigating the German language. They have found evidence that different types of pitch accent signal different degrees of givenness. Accordingly, the term givenness is relevant as it accounts for varying degrees of activeness in the sense of Chafe (1976; 1987). Lambrecht (1994), on the other hand, formulates activeness as a binary category.

In his account, Chafe (1976: 38ff) also discusses definiteness among a number of linguistic phenomena that appears to intersect with givenness. In English, the definite noun phrase encodes referents that have already been mentioned in the discourse and the speaker assumes that the addressee knows them. Chafe (1976: 39) emphasizes,

though, that if a speaker believes that the listener is familiar with the referent the speaker has in mind, the speaker would encode it with a definite noun phrase without prior mention of the referent. For example, if the speaker has a dog and the listener knows the dog, then without any prior mentioning of the dog, the speaker might utter: "*Sorry I'm late, I had to feed* ***the dog***." In order to capture the knowledge that the speech participants may have about the circumstances, Lambrecht (1994) adopts the term *identifiability*. If the listener knows or is familiar with the referent mentioned in an utterance, then the referent is identifiable. If the referent is unknown to the listener, then it is unidentifiable. As for the grammatical means of identifiability in Lambrecht (1994), the situation is similar to activation: knowledge is continuous, whereas the linguistic category is a binary choice between identifiable and non-identifiable (Lambrecht, 1994: 84). In English and in many other languages, the grammatical correlate for the category of identifiability is definiteness (Lambrecht 1994: 79). Other possible formal means can be word order and presence or absence of numeral or morphological case marking (Lambrecht 1994: 79).

Another interesting characteristic about expressions that encode referents as identifiable is that the referent is not necessarily presupposed to exist. In a text-internal context the speaker can refer to *The King of France* and ignore the fact that the King of France does not exist, but the speaker assumes (because he probably has established it before or is going to establish *the king* in the following context) that the addressee is able to have or to create a mental referent for this expression (Lambrecht 1994: 76). This can be accounted for as *pragmatic accommodation* (Lambrecht, 1994: 66) – a phenomenon in which the speaker creates a new presuppositional situation by "using merely an expression that requires this situation". Therefore, an expression encoding a referent as active or identifiable does not necessarily refer to an activated or identifiable referent in the discourse.

The idea of pragmatic accommodation fits with the aspect of information structure recognized by Ellen F. Prince (1981). The problem of the theories of information structure is that they appear to capture information as something that speakers/writers and listeners/readers are able to assess on the basis of linguistic form in the discourse (Prince, 1981: 233). However, there is no one-to-one mapping between the form and the information that exists independent of the minds of the speakers – 'objectively' (Prince, 1981: 233). Therefore, Prince suggests that grammatical devices

used for expressing information structure should be interpreted as a "set of instructions from a speaker to a hearer on how to construct a particular discourse-model" (Prince 1981: 235).

1.1.3. Interim Summary

The current study adopts the concepts of focus (Halliday, 1967ab; Lambrecht, 1994; Gundel, 1999), activeness (Chafe, 1976; 1987; Lambrecht, 1994), identifiability (Lambrecht, 1994) and givenness (Chafe, 1976; Baumann, 2006) in order to deal with the complex interaction between word order and intonation in the following chapters.

Focus is that part of an utterance which the speaker decides to present as new information (Halliday, 1967ab; Lambrecht, 1994). By employing certain linguistic devices, the speaker makes an effort to draw the listeners' attention to some information (Gundel, 1999). This information does not necessarily have to be new to the discourse. Focus can also contain old information (Halliday, 1967ab; Lambrecht, 1994; Gundel, 1999). A category of activeness encodes the status of the referent in the mind of the speech participants during a speech event (Chafe, 1976; 1987; Lambrecht, 1994). A referent might become activated inferentially, textually and situationally (Lambrecht, 1994). Identifiability as an information-structural category accounts for the referents that are either known or unknown to both speech participants (Lambrecht, 1994). Similarly to activeness, givenness encodes the informational status of a referent in a discourse, but is a category that as more degress than just given and new (Chafe, 1976; 1987; Baumann, 2006; Baumann and Grice, 2006; Baumann and Riester, 2012).

Just as a linguist is not able to estimate what is in the minds of speech participants, neither are speakers able to estimate entirely what is active or identifiable in the discourse. They construct their utterances based on predictions about the discourse and the addressee's knowledge about the discourse (Chafe, 1976; Lambrecht, 1994). Therefore, the speaker decides whether to represent some information as given or as focus (Halliday, 1967a; Lambrecht, 1994) and it is on the addressee to accept it or not (Lambrecht 1994: 103).

Thus, the speech participants have a possibility to negotiate the information-structural status of a sentence constituent. This possibility, however, implies unambiguous linguistic devices that are capable of marking the information status

intended by the speaker. The main linguistic devices of focus are accent and word order (Lambrecht, 1994; Gundel, 1999). The possible linguistic devices of activeness are lexical encoding (full noun *vs.* pronoun) and accent (Chafe, 1976; 1987; Lambrecht, 1994; Gundel, 1999). The most familiar grammatical means for identifiability is definiteness, but there might be other means such as word order or morphological markers implemented for it (Chafe, 1976; Lambrecht, 1994). Following Prince (1981), the formal means of information structure are seen as directions on how the addressee is expected to interpret the informational status of the referents involved in a discourse.

The study relies largely on the theory of information structure put forward by Lambrecht (1994). Similarly to Krifka (2007), he emphasizes that linguistic devices of information structure and the information-structural categories are to be defined independent of each other. In addition, the definition of focus as a speaker's presentational choice (Lambrecht, 1994) is compatible with the formalization of focus put forward by Rooth (1992) and strongly advocated by Krifka (2007). The three-way articulation of information structure (focus, identification and activation) adopted from Lambrecht help to gain first insights about the interaction between word order, information structure and intonation for a language that is yet clearly insufficiently investigated in terms of intonation and information structure. The generalizations derived by the juxtaposition of the theory and the experimental results presented in the thesis are plausible also independent of the theory in Lambrecht (1994). Next, the two formal means of information structure will be discussed: first, sentence accent and second, word order.

1.2. Sentence accent

1.2.1. Pitch prominence

Lambrecht (1994: 210–214) refers to 'accent', Halliday (1967a: 203) to 'tonic' in terms of pitch movement and Chafe (1976: 31) to high pitch and strong 'stress'. All these observations refer to some kind of *prosodic prominence* that an expression in focus appears to have in English in relation to other words in a sentence. The question arises as to how this abstract prominence, which a number of theoretical accounts of information structure relies on, is manifested in a spoken language?

The physical/phonetic basis of prominence has been shown to have higher fundamental frequency (F0), longer segment durations, greater intensity and increased spectral emphasis (Fry, 1955; 1985; Gussenhoven et al., 1997; Rietveld & Gussenhoven, 1985; Terken, 1991; Beckman, 1986; Beckman & Edwards, 1994; Turk & White, 1999; Cambier-Langeveld & Turk, 1999; Kochanski et al., 2005; Sluijter & Van Heuven, 1996). Phonologically, there are two domains at which prominence occurs. *Word-level prominence* (also lexical stress) refers to a syllable in a word that is more prominent than other syllables of the same word. *Phrasal prominence* (also accent) refers to one or more words that are more prominent than other words in the same phrase. In other words, prominenenc can determined for a word as *word-level prominence* and for a phrase as *utterance-level prominence* (e.g. Lehiste, 1970). Obviously, information-structural category focus is concerned with phrasal prominence.

Studies have shown that the main acoustic correlate of phrasal prominence is F0 in English (Cooper et al., 1985; Eady & Cooper, 1986; Campbell & Beckman, 1997; Swerts et al., 2001; Breen et al., 2010) and in German (Baumann et al., 2006; Féry & Kügler, 2008). There are some studies (Sluijter & Van Heuven, 1996; Campbell & Beckman, 1997; Suomi et al., 2003) that aim to establish that the acoustic correlates are different for word-level *vs.* sentence-level prominence. Agaath M.C. Sluijter and Vincent J. van Heuven (1996) find that for Dutch the perception of word-level prominence relies mainly on vowel duration and intensity, but not on F0. For Finnish, Kari Suomi, Juhani Toivanen and Riikka Ylitalo (2003) find that the sentence-level prominence is clearly realized with higher F0, but not the word-level prominence. Nick Campbell and Mary Beckman (1997) replicated the study by Sluijter and Van Heuven (1996) for English but did not find such a clear difference between the acoustic correlates of word-level and sentence-level prominence. In fact, the acoustic correlate of phrasal prominence – higher F0 – is almost always accompanied by greater duration and intensity (Beckman & Edwards, 1994; Breen et al., 2010).

A theoretical solution for interpreting the acoustic prominence at word and phrase-level is offered by an abstract *prosodic or metrical structure* (e.g. Selkirk, 1984; 1995; Nespor & Vogel, 1986; Ladd, 2008) by which utterances are always made up of syllables, feet, phonological words, phrases and intonational phrases. It is a hierarchical structure with higher-level elements depending on lower-level elements, in the sense that the words are composed of feet and feet of syllables (see e.g. Gussenhoven,

2004: 123–125 or Ladd, 2008: 55–56). Abstract levels of strong and weak forms in a word and of the words in a phrase represent prominence relations between syllables. A prominence relation that is specified at some lower level (strong *vs*. weak syllables) becomes less specified at a higher level of the structure (strong *vs.* weak words). In example by Ladd (2008: 55), syllables in the word *baby* are specified for stress (*ba-* is strong and *-by* weak). If *baby* forms a part of a compound word *babysitter,* then different parts of the compounds are specified for stress (*ba-* is stronger than *si-*). However, prominence value for each syllable in a compound is left unspecified. At the phrasal level, the prominence value of each word is similarly left unspecified. The acoustic correlates mentioned above (duration, F0, intensity, spectral emphasis) are interpreted within this prosodic structure.

Hence, when Lambrecht (1994: 210–214) refers to 'accent', it is likely that he means abstract sentence-level prominence, whereas Halliday (1967a: 203) and Chafe (1976: 31) – by concerning a pitch movement or high pitch – refer to one of the (main) acoustic correlates of phrasal prominence specifically relevant for English and German.

Phrasal prominence conveyed by F0 is also known as *intonational* or *pitch prominence* (see e.g. Terken & Hirschberg, 1994; Baumann & Grice, 2006). Intonation and pitch are strongly related to F0. A speech sound as a quasi-periodic sound wave contains multiple time-aligned low- and high-frequency vibrations in an elastic medium (air, membrane in the microphone). A common denominator, also the lowest frequency of all those vibrations, is called *fundamental frequency* (F0). F0 corresponds roughly to the frequency of glottal pulses that are physiologically the source of sound waves. The main way to change F0 is by varying vocal fold tension: the constriction of intrinsic laryngeal muscles raises the F0 and the relaxing lowers it (Shipp & McGlone, 1971; Collier, 1974; Honda, 2004). Also, the subglottal air pressure (Collier, 1974) and the height of the glottis (Honda, 2004) affect the F0. F0 is perceived as *pitch* and at the linguistic level F0 is called *intonation* (or *lexical tone*). Human perception of F0 is non-linear and, therefore, it is advised to measure and observe F0 excursions on a *logarithmic semitone scale* (Cohen et al. 1982: 264; t' Hart et al., 1990: 23). All three terms – F0, pitch and intonation – are used quite interchangeably in the following chapters.

In a phrase, pitch prominence (e.g. of a referring expression) is perceived relatively to the pitch of other words in the same phrase. See Figure 1.1 that depicts F0 (Hz) as a function of time (ms).

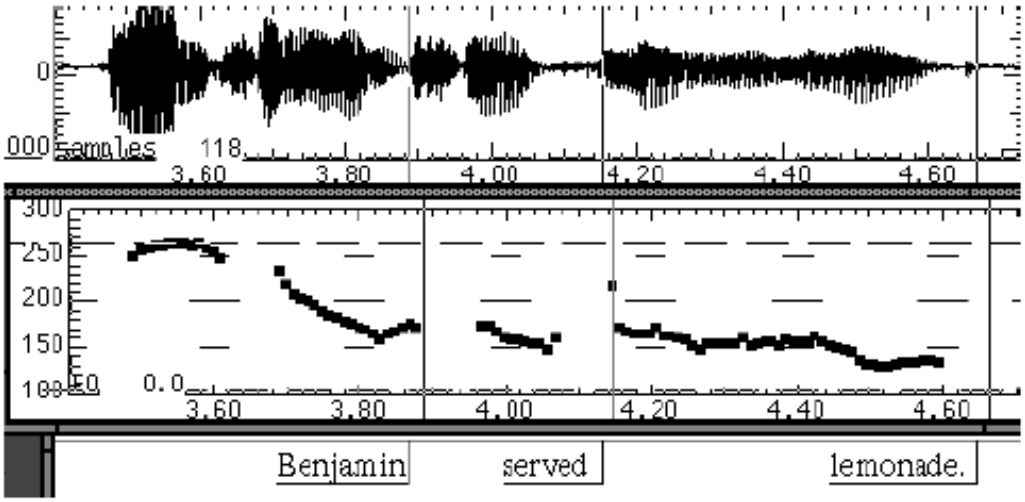

Figure 1.1. Pitch track of the sentence *Benjamin served lemonade* with the focus on *Benjamin,* frequency (Hz) on the y-axis, time (ms) on the x-axis (Jannedy, 2002: 13).

In Figure 1.1, F0 is high (about 230 Hz) at the beginning of the utterance and low (about 150 Hz) in the remainder of the utterance. High F0 is perceived as prominent only in relation to the following low F0 that does not show any movements. Similarly, high F0 (about 200 Hz) on the words *Benjamin* and *served* in Figure 1.2 is perceived as non-prominent in relation to the fall from high F0 (about 210 Hz) to low F0 (about 150 Hz) on the word *lemonade*.

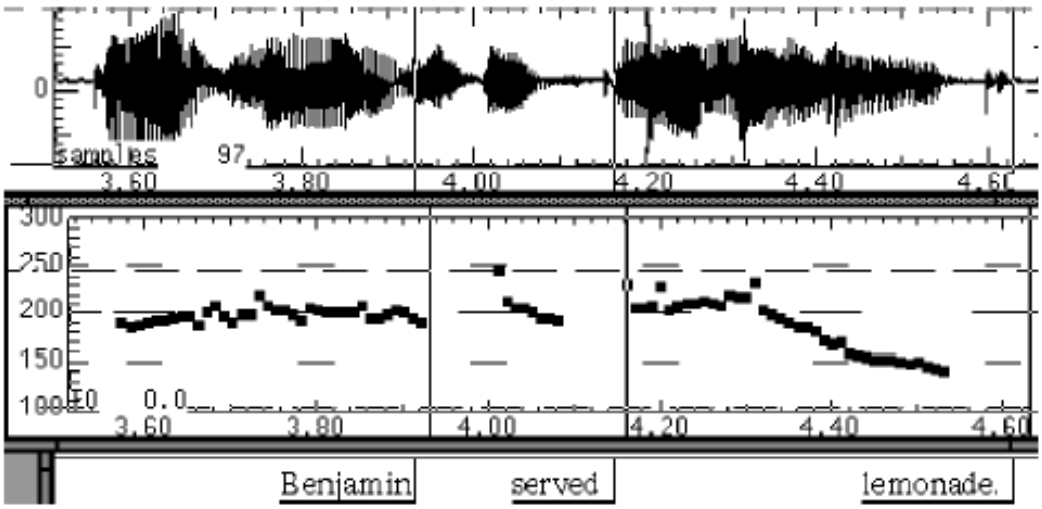

Figure 1.2. Pitch track of the sentence *Benjamin served lemonade* with *lemonade* in focus, frequency (Hz) on the y-axis, time (ms) on the x-axis (Jannedy, 2002: 13).

Examples of pitch tracks in Figures 1.1 and 1.2 illustrate another relevant phenomenon called *accent shift* (Ladd, 2008). Observe in both figures that pitch prominence changes its location from the beginning to the end of the phrase. Intonationally speaking, the word *Benjamin* in Figure 1.1 is *pitch accented* and the phrase *served lemonade deaccented*, whereas the word *lemonade* in Figure 1.2 carries *pitch accent* and the phrase *Benjamin served* is *deaccented*. In respect to focus, the utterance in Figure 1.1 might originate from the context in (1.6) and the utterance in Figure (1.2) from that in (1.7).

(1.6)

a. Who served the lemonade?

b. BENJAMIN served lemonade.

(1.7)

a. What did Benjamin serve?

b. Benjamin served LEMONADE.

Deaccentuation in a phrase has been reported to be a relevant cue for old information (givenness, activated referents) in intonation languages (e.g. English and German). Many studies report that expressions referring to given information are deaccented (to name just a few: Brown, 1983; Terken & Hirschberg, 1994;

Pierrehumbert & Hirschberg, 1990; Baumann & Hadelich (2003); Cruttenden, 2006). Xu and Xu (2005) demonstrate for English that deaccentuation in postfocus position consistently causes F0 to be flat (as seen in Figures 1.1 and 1.2 above).

In connection to accent as a focus-marking device, the term *focus domain* needs to be addressed. Focus domain is the syntactic domain (a noun phrase, a verb phrase or a whole sentence) that consists of information presented as new to the discourse. In other words, focus domain expresses focus. There are instances of indirect correspondence between accent placement and a focus domain, which might cause ambiguity in focus interpretation, as in the example (1.8) (the capital letters hereafter indicate the accented constituent).

(1.8)

a. What does he teach? – He teaches LINGUISTICS.
b. What does he do? – He teaches LINGUISTICS.

The phenomenon in (1.8) is often called *focus projection* in syntactic accounts of sentence accent (Gussenhoven, 1983; Selkirk, 1984; 1995). In both sentences in (1.8), the accent is on the last word of the sentence. In (1.8a), focus domain is the noun phrase *linguistics*. In (1.8b), however, it is the whole sentence. Since accent does not mark the focus constituent directly, it is impossible to determine the focus of the utterance without additional context.

Another case of indirect correspondence between accent and focus domain that is connected to the differences in syntactic functions of the sentence constituents is shown in examples in (1.9).

(1.9)

a. I heard your motorcycle broke down? - My CAR broke down.
b. What happened? - My CAR broke down. (Lambrecht 1994: 223)

In the examples in (1.9), it can be seen that different focus domains – the whole utterance in (1.9b) and the noun phrase in utterance (1.9a) – are both marked with an accent on the noun phrase. In this phenomenon of argument-predicate asymmetry, the verbs next to the nominal constituents (like sentence arguments) are usually deaccented (see Gussenhoven, 1983, 1992 for further details).

Examples (1.8) and (1.9) suggest that in addition to a simple correspondence between focus and accent, there appear to be additional preferences for the accent placement at the syntax-phonology interface. However, examples (1.8) and (1.9) should not pose any further problems, if focus is defined as a pragmatic relation that cannot be always detected on a particular linguistic form of a sentence constituent (Lambrecht 1994: 213).

Another distinction considering focus domain is made in different theories of focus (e.g. Krifka, 2007; Gussenhoven, 2007; Ladd, 2008). Namely, focus types *broad* and *narrow* refer to the width of the focus domain. A focus domain consisting of one word (1.9a) is called narrow focus, a focus domain consisting of a verb phrase (1.9b) or some larger unit like a whole utterance is called broad focus.

1.2.2. Intonational categories

The object of the study of intonation are changes from high to low F0 (as seen in Figure 1.1 and 1.2) or from low to high F0, that are perceivable and linguistically functional. It was first demonstrated in the studies of English and Swedish intonation (Pierrehumbert, 1980; Bruce, 1977) that the F0 minima and maxima tend to be *associated* with the stressed vowel of a word. The F0 contour can be therefore segmented into tonal categories based on F0 minima or F0 maxima that are associated with the word's most prominent syllable. Crucial elements of intonational phonology are *pitch accents* and so-called *edge tones* (Ladd, 2008). The theory of tonal categories is called the autosegmental-metrical (AM) theory of intonation.

Pitch accent is a conspicuous F0 event in a continuous F0 contour that is associated with the stressed syllable of a prominent word in a phrase. An important property of pitch accent is that it is a F0 curve that is perceived as prominent (Ladd 2008: 8). Abstract phonological levels low and high (Ladd, 2008: 62ff) combine into simple tones – high (H) or low (L) – and into complex ones (Pierrehumbert and Hirschberg, 1990). The high (H) or low (L) tone of a stressed syllable rhyme is referred to as a *starred tone* (H*, L*). The starred tone may be followed or preceded by a perceptually distinctive *trailing* or *leading* tone, respectively, and is described in terms of its tonal level (H or L). Tones that only consist of a starred tone are referred to as *monotonal pitch accents* (e.g. H*, L*), whereas tones containing either a leading or a

trailing tone are called *bitonal* pitch accents that reflect F0 rises (L+H*, L*+H) or F0 falls (H+L*, H*+L). The bitonal marking of a pitch event reflects the fact that the stressed vowel may be characterized by perceptually salient F0 transition instead of a level tone. A bitonal pitch accent with a leading tone (L+H*, H+L*) has H or L target that is reached relatively late within the stressed vowel (at the end of it or already after it), whereas the bitonal pitch accent with the trailing tone (H*+L, L*+H) has H or L target that is reached relatively early in the stressed vowel (at the beginning of it or even before it). The location of H and L relative to the stressed vowel is accounted for in the theory of *alignment of tonal target* (Arvanti et al., 1998 for Greek; A; Ladd et al., 1999 for English and Ladd et al., 2000 for Dutch).

The intonational phrase may be marked with an additional tone that is called *boundary tone* (Pierrehumbert, 1980; Pierrehumbert & Hirschberg, 1990). Mary E. Beckman and Janet B. Pierrehumbert (1986) distinguish between two types of intonational phrases – *intermediate phrase* and *intonational phrase.* The intermediate phrase may end with a *phrase accent* (Pierrehumbert, 1980). Phrase accent associates with stress and is a *stress-seeking* tone, whereas boundary tones just mark the edges of a phrase and are *edge-seeking* (Gussenhoven 2004: 140). Boundary tones and phrase accents are both called *edge tones* in Ladd (2008). An edge tone is similar to a pitch accent, either at low (L) or high (H) level (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986; Pierrehumbert & Hirschberg, 1990; Ladd, 2008; Gussenhoven, 2004). The intermediate phrase boundary is transcribed with a hyphen (H-, L-) and the intonational phrase boundary with a percent sign (H%, L%; Beckman & Pierrehumbert, 1986).

In brief, an intonational phrase can be seen as a string of tones – of pitch accents and of different kinds of edge tones. The well-known application of autosegmental-metrical (AM) theory is ToBI ('Tones and Break Indices') transcription (Beckman & Elam, 1997) and there are conventions available for a number of different languages (see Jun, 2006 for an overview). In example (1.10), a demonstration of one sort of a ToBI transcription for English can be seen.

(1.10)

H* H* LH%

Could I have the bill please. (Ladd 2008: 114)

Janet B. Pierrehumbert and Julia Hirschberg (1990: 286) claim that different types of pitch accent correspond to different types of information status such as 'mutually believed', 'inferable' or 'identifiable'. The H* pitch accent conveys that the referring expression is "to be treated as 'new' in the discourse" (Pierrehumbert & Hirschberg, 1990: 289). The L* accent occurs on active (salient) referents in a discourse (Pierrehumbert & Hirschberg, 1990: 291). The L*H conveys uncertainty, incredulity and a lack of speaker commitment (Pierrehumbert & Hirschberg, 1990: 294), whereas the H*L establishes inference relationships between items in subsequent utterances. In their generalization (Pierrehumbert & Hirschberg, 1990: 301), pitch accents with a low leading tone (L*+H, L+H*) signal the entities already active in the discourse or entities that do not need to be added into the hearer's mutual belief anymore, whereas pitch accents (H*+L, H+L*) with a high leading tone transmit new information that must be added to the mutual belief of the speech participants.

Similar phenomena have been observed for German where the information structure of a sentence may be determined by the alignment of high tones. An early-aligned F0 peak (H*+L) supports the interpretation of an utterance as established information, or even as a committed or sarcastic statement (Kohler, 1987b; Grice et al., 2005). A medial peak alignment (H*), on the other hand, supports the interpretation of an utterance as new information (Kohler, 1987b; Grice et al., 2005) and a late peak alignment (L*+H) supports the interpretation of an utterance as an incredible piece of information (as new information that is contrasted to the previous informational state; Kohler, 1987b; Grice et al., 2005;). Recent studies (Baumann & Grice, 2006; Baumann & Hadelich, 2003) establish on the continuous scale of activeness that the absence of accent signals active concepts, the H+L* accent the semi-active information and the H* accent the inactive information.

1.2.3. Prenuclear and nuclear pitch accents

Pitch accents and edge tones can be studied independently of their pragmatic functions or information structure since they are independent grammatical elements of intonational phonology (Ladd, 2008). However, intonational phonology and information structure intersect with each other in the concept of *nuclear accent* ('tonic' in Halliday 1967b; Beckman & Edwards, 1994; Ladd, 2008;). In intonational phonology, the last accent of a prosodic phrase (either intermediate or intonational) is

called *nuclear pitch accent* (Beckman & Edwards, 1994; Ladd, 2008). All pitch accents preceding it are called *prenuclear pitch accents* (Beckman & Edwards, 1994; Ladd, 2008). Note that the definitions of prenuclear and nuclear pitch accent involve just their position in a string of pitch accents. What gives nuclear pitch accent its special status is its primariness: it is basically the only obligatory pitch accent in a phrase (Halliday, 1967ab; Ladd, 1986; Ladd, 2008: 133).

The primariness of nuclear pitch accent arises from two reasons: first, from the structure of a prosodic phrase and second, from pragmatic reasons. The structural needs of a prosodic phrase are quite obvious: it is hard to imagine an intonational phrase without a pitch accent, basically without any prominent F0 curve. In English, pitch accent in a way constitutes a prosodic phrase, but this does not necessarily have to be so for other languages. A prosodic phrase can also be defined by its boundaries, e.g. by boundary tones or phrase accents. For example, Anja Arnhold (2014) argues that Finnish is a phrase accent language. This fact, however, does not change the main assumption: a prosodic phrase needs to have some defining F0 movement (Ladd, 1986); in intonation languages (e.g. in English) it just happens to be so that it is pitch accent.

Pragmatically, if the grammatical device of focus appears to be the accent, then it is a nuclear accent: "only primary or nuclear accents are relevant to signalling focus", whereas "secondary accents are distributed according to other criteria" (Ladd, 2008: 266). If there are no other pragmatic factors present, then nuclear pitch accent is preceded by prenuclear pitch accents in a longer phrase (see Ladd, 2008: 159ff). Prenuclear pitch accent is optional and the listeners perceive them differently with respect to their prominence status (Cutler & Foss, 1977; Ayers, 1996; see also Ladd's discussion of ToBI labeling examples in 2008: 261). Therefore, prenuclear pitch accents are unlikely to play a crucial role in the pragmatic interpretation of a phrase and it follows that not every conspicuous F0 movement results in pitch prominence.

Phonologically, the same types of pitch accents can occur nuclearly as well as prenuclearly, although the position might cause some phonetic effects such as early peak alignment of nuclear pitch accents (Silverman & Pierrehumbert, 1990). In terms of phonetics, however, nuclear pitch accents do not differ from prenuclear pitch accents (Silverman & Pierrehumbert, 1990; Ladd, 2008). The F0 excursions occurring nuclearly are the same size or even smaller than the excursions occurring prenuclearly.

As Ladd notes: “The last accent is often low in pitch and in overall intensity” (2008: 257). What matters is that accent is present in a prosodic phrase (Ladd, 2008: 259), and its presence may well be defined by acoustic features other than F0. A nuclear pitch accent that is missing conspicuous F0 movement is realized probably with other acoustic features that usually would be attributed to the word-level prominence at first sight. Crucially, prominence is defined relationally within abstract *metrical structure* where the lower level (word-level) prominence can be left unspecified and, therefore, the acoustic features that are reported to characterize word-level prominence (like established by Sluijter & Heuven, 1996) are interpreted rather at the phrase level. For this reason, the theory of relational prominence is important here. Consequently, many additional accents may be present in a phrase, but as long as one of the accents is for some reason (syntactic, lower scaling of other pitch accents in a phrase, the final position in a string) perceptionally more prominent, it is a nuclear accent.

1.2.4. Interim summary

This subsection introduced the physical/phonetic background of prominence underlying many theoretical accounts of information structure. In English, pitch accents and accent shift are strongly involved in prominence perception and marking the informational status of referents in a discourse. On the other hand, the discussion of prenuclear and nuclear pitch accent demonstrated that not all conspicuous F0 movements contribute to a pitch prominence (prenuclear) or at least that they are not always pragmatically motivated.

We also discussed the intonational categories of pitch accents and boundary tones in order to show that different types of pitch accent can convey information about the informational status of a referent in a discourse. Therefore, it is not just the presence or absence of prosodic prominence that is relevant for the study of formal means of information structure; other intonational phenomena may also play a role. The next section discusses another type of linguistic device marking information status: word order.

1.3. Word order

1.3.1. Free word order

There are often various possibilities to arrange words into a meaningful utterance. Consider examples (a) and (b) in (1.11).

(1.11)

a. Here comes the CAT.
b. And here the cat COMES. (Lambrecht 1994: 39, 41)

In both examples, the constituents of the sentences in (1.11) are a sentence modifier or an adverbial *here* (X[1]), a verb (V) *comes* and a subject (S) noun phrase *the cat*. Observe that the order of constituents is different in the examples: in example (a) it is XVS, whereas in (b) it is XSV. Subject noun phrase like *the cat* in example (1.11) or the object (O) noun phrase like *John* in the sentence *Mary loves John* are called *arguments* in a syntactic theory. They can be seen as the main constituents of a sentence, because the valency of the verb (the number of nominal constituents that the verb can minimally take)[2] determines their presence in a sentence, whereas the presence of adverbials is free. Since nominal constituents can have different grammatical functions (e.g. subject or object), their order is relevant for the interpretation of grammatical relations in English, as in examples in (1.12).

(1.12)

a. My sister meets the boss tonight.
b. The boss meets my sister tonight.

In example (1.12a), the noun phrase *my sister* is the subject and as such precedes the verb *meet* (word order SV). However, the noun phrase *my sister* is the object in (1.12b) and follows the verb (word order VO). Thus, English appears to have a main constituent order SVO that is crucial for interpreting the grammatical relations between the referring expressions.

---

[1] In the theoretical literature X usually stands for an optional sentence constituent that is not the object or subject noun phrase.

[2] A transitive verb in English like *to meet* can take at least two nominal constituents (*My sister meets the boss*) whereas intransitive verb like *to sing* can take at least one nominal constituent (*My sister sings*).

Languages differ in their flexibility of ordering arguments such as subject or object noun phrases in relation to the verb (Siewierska, 1998). English or French, in which word order encodes the grammatical case (subject *vs.* object), are called *configurational languages* (Baker, 2001). Other languages do not seem to encode grammatical relations in word order and are, therefore, free to vary the combinations of argument-verb order: SVO, OVS, SOV, OSV, VSO, VOS. They are, therefore, called *non-configurational languages* (Baker, 2001). See, for instance, the examples from Warlpiri language in (1.13) (bold emphasizes the word that varies its position).

(1.13)

a. **Kurdu-ngku** ka-ju nya-nyi ngaju.
child-ERG PRES-1Sg.O see-NPST I (ABS)
'The child sees me.'

b. **Kurdu-ngku** ka-ju ngaju nya-nyi.
child-ERG PRES-1Sg.O I (ABS) see-NPST
'The child sees me.'

c. Nya-nyi ka-ju **kurdu-ngku** ngaju.
see-NPST PRES-1Sg.O child-ERG I (ABS)
'The child sees me.'

d. Ngaju ka-ju nya-nyi **kurdu-ngku**.
I (ABS) PRES-1Sg.O see-NPST child-ERG
'The child sees me.'

e. Ngaju ka-ju **kurdu-ngku** nya-nyi.
I (ABS) PRES-1Sg.O child-ERG see-NPST
'The child sees me.'

f. Nya-nyi ka-ju ngaju **kurdu-ngku**.
see-NPST PRES-1Sg.O I (ABS) child-ERG
'The child sees me.' (Simpson 1983: 140)

Observe that the nominal constituent *kurdungku* is shifted into a different sentence position in the examples (a) – (f) in (1.13) without changing the meaning or perspective of the utterance: in all versions of the utterance it is the child who sees the speaker and not the other way around. The phenomenon by which the sentence constituents can switch their location in a sentence without a change in grammatical interpretation (as in the example of Warlpiri in (1.13)) is called *free word order*.

Anna Siewierska (1998: 504) refers to languages with and without free word order as *rigid* and *highly flexible word order* languages. Languages differ in the degrees of flexibility in word order. Therefore, two additional categories for intermediate free word orders are *restricted word order variation* and *flexible word order* (Siewierska, 1998: 504). This terminology makes no reference to syntactic configurations and accounts for word order variation in terms of number of possible word order permutations: a language that employs all six permutations of subject, object and verb is highly flexible, whereas a language that has only one word order variant is a language with restricted word order variation.

Flexible word order correlates strongly with morphological marking of syntactic functions (Jelinek, 1984; Steele, 1978; Siewierska, 1998): observe in the example of (1.13) that the noun phrase 'kurdungku' consists of a suffix *-ngku* that marks the grammatical function of the subject. Siewierska (1998: 507f) finds that 81% of languages with no overt morphological marking of functions (or semantic roles) are rigid or restricted word order languages, whereas languages with morphological marking are highly probable (80% or more) to employ more than three word order variants. However, like Siewierska (1998) and Primus (2001) discuss, the morphological marking does not incur the flexibility in word order. For example, Icelandic (Siewirska, 1998) has rigid word order, although the language applies verb-agreement and morphological marking.

Pragmatic factors influence the order of the constituents in a number of free word order languages (Mithun, 1987; É. Kiss, 1995). For example, expression in focus occurs in a certain position in relation to the verb and consequently, only certain word order might be appropriate in a particular discourse. Therefore, free word order only refers to word order flexibility from a syntactic perspective (specifically from a phrase structure grammatical perspective (Chomsky, 1957)). The next section deals with some pragmatic factors that are known to influence word order.

1.3.2. Information structure and free word order

There are two ways to account for information-structural effects in a free word order language. The first account is rooted in syntactic analysis and is known as *discourse-configurational* account of free word order. In a discourse-configurational account of word order, the information-structural category topic or focus is encoded in the

syntactic structure (in the syntactic tree) similarly to the grammatical relations of subject and object noun phrase in a configurational language (É. Kiss, 1995; 2001; Rizzi, 1997). This means that the topic or focus needs to be located in a structural position related to the verb position in an utterance; see an example from Hungarian in (1.14).

(1.14)

Q1: What did Melanie eat?

A1:	Melánia	**mandarint**	evett.
	Melanie	tangerine.ACC	ate

Q2: Who ate the tangerine?

A2:	**Melánia**	evett	mandarint.
	Melanie	ate	tangerine.ACC

In example (1.14), the nominal constituent in focus (*mandarint*, *Melánia*) is located before the verb *evett* in both examples. The shift of the constituent in focus to the preverbal position is obligatory in Hungarian; this in turn is a crucial characteristic for defining a language as discourse-configurational: a language is discourse-configurational, if it has either *structural topic* or *structural focus* (É. Kiss, 1995: 6). There are numerous languages that are reported to have structural focus or structural topic (see É Kiss, 2001 for an overview).

Another way to account for the information-structural effects due to word order is based on construction grammar (Fillmore, 1985). This account takes different word orders (e.g. SVO, OVS, SOV, OSV, VSO, VOS) that are possible in a language as constructions that come with specific semantic and pragmatic properties (Välimaa-Blum, 1988; 1993; Vilkuna, 1998). The Finnish existential clause in (1.15) serves as an example (Välimaa-Blum, 1988).

(1.15)

a.	**Kissa**	nukkuu	sängyssä.
	cat.NOM	sleep.3Sg	bed.INE
	'The cat sleeps in the bed.'		
b.	Sängyssa	nukkuu	**kissa**.
	bed.INE	sleep.3Sg	cat.NOM
	'The cat sleeps in the bed.'		

c. **Kissa** sängyssa nukkuu.
cat.NOM bed.INE sleep.3Sg
'The cat sleeps in the bed.'

d. Sängyssä **kissa** nukkuu.
bed.INE cat.NOM sleep.3Sg
'The cat sleeps in the bed.'

e. Nukkuu **kissa** sängyssä.
sleep.3Sg cat.NOM bed.INE
'The cat sleeps in the bed.'

f. Nukkuu sängyssä **kissa**.
sleep.3Sg bed.INE cat.NOM
'The cat sleeps in the bed.'

The nominal constituent *sängyssä* is an adverb (X), *nukkuu* a verb (V) and *kissa* a subject (S) noun phrase. The examples of (a) – (f) in (1.15) exhibit six word order variants: SVX, XVS, SXV, XSV, VSX, VXS. Riitta M. Välimaa-Blum (1988: 74) finds in her analysis that the first two orders (SVX and XVS) – subject-initial and subject-final order – are neutral in respect to information structure; the medial two orders in (1.15) (SXV and XSV) are the orders where the sentence-initial constituent – either *kissa* (S) or *sängyssä* (X) is contrasted to some elements in a discourse; and in the last two orders (VXS and VXS), the whole proposition is 'emphatic' according to Välimaa-Blum (1988). Compatibly, Maria Vilkuna (1998: 193) notes that the verb-initial orders are either very strong confirmations of the truth of the proposition or the whole proposition is contradicted to a discourse.

Välimaa-Blum (1988: 62) observes for Finnish that the identifiability or specificity of a referent can be reflected in the position of the referring expression in a sentence. In the examples in (1.16) the word *ukko* ('a man') is at end of the sentence in (a) and at the beginning of the sentence in (b).

(1.16)

a. Tuvassa on ukko.
cottage-INE is man-NOM
'There is a man in the cottage.'

b. UKKO on tuvassa.
man-NOM is cottage-INE
'The man is in the cottage.'

The semantic effect of (1.16a) is that the man standing in the cottage is unidentifiable: he has just been introduced into the discourse and the speaker is probably going to say something more about him to the addressee. In (1.16b), all speech participants already know *the man*. A possible context could be that the speech participants both expected the man, but one of them noticed him in another room and uttered the sentence to his addressee. Kaiser and Trueswell (2004) have demonstrated in an eye-tracking study for Finnish that in the sentence with OVS word order, the listeners do interpret the subject noun phrase as new to the discourse.

Elena Titov (2012) observes a similar phenomenon while speaking about *referentiality*[3] in Russian. Titov (2012) shows for Russian that if there is no specific context and the expressions are equal in terms of focus (they might both stand in focus), then the non-referential object follows a referential object. Observe examples in (1.17).

(1.17)

a. Ivan peredal špiona **agèntu**
Ivan handed spy.ACC agent.DAT
'Ivan handed the/a spy to the/an agent.'

b. Ivan peredal **agentu** špiòna
Ivan handed agent.DAT spy.ACC
'Ivan handed a spy to the agent.'

In example (1.17) *a spy* is a direct object and *to the agent* an indirect object. In (a) the word *agent* is at the end of the sentence, whereas in (b) it is in the middle of the sentence. The sentence (1.17b) is appropriate in a context where *the agent* is identifiable and locatable (specific) for an addressee, whereas the sentence (1.17a) does not evoke such an implication.

Thus, the examples of Finnish and Russian demonstrate that in addition to topic or focus of an utterance, word order might also encode other information-structural

[3] By referentiality Titov means the pragmatic category for expressions that have or do not have a counterpart in the text-external world (see Abbott, 2006 for further definitions and discussion)

phenomena like identifiability. There is additional data presented by Gregory Ward and Betty Birner (2004) for English and by Thomas Weskott, Robin Hörnig, Gisbert Fanselow and Reinhold Kliegl (2011) for German that could be reinterpreted in terms of activation (Lambrecht, 1994; Chafe, 1987).

Ward and Birner (2004) show that English is not always obliged to use prosodic marking of focus or old information/topic but can also optionally use different structural means like preposing, postposing, argument reversal, passivization etc. for highlighting information structure of an utterance. Consider the example of preposing in (1.18).

(1.18)

Q: Can I get a bagel?

A: No, sorry. We are out of bagels. A bran muffin I can give you.

Observe in (1.18a) that the object noun phrase *a bran muffin* stands in a 'non-canonical' position for an object noun phrase in English – at the beginning of the sentence. Ward and Birner (2004: 159) argue that the sentence-initial position for the expression *a bran muffin* is licensed by the preceding expression *bagel* that evokes a *poset* – a partially ordered relation. A poset consists of items that are in some (semantic) relation to each other. A semantic relation, for example, could be a relation of type and subtype (*pie* and *desserts*), greater than (*five* and *six*) or some sort of inclusion (*oranges* and *apples*). However, another way to look at these relations is in terms of *activation*: the mentioned item activates all other items that belong to the same set (for example, mentioning of oranges activates apples in a discourse (in the addressee's mind)). Lambrecht calls this type of activation *textually accessible* (1998: 100); for Chafe (1987), the referent is *semi-active*.

Weskott et al. (2011) explain appropriateness of OVS word order in German also with the theory of poset. In the German example (1.19), different word orders in (a) and (b) are two possible continuations of context (C).

(1.19)

C. Peter	hat	den	**Wagen**	gewaschen.
Peter	has	the.ACC	car	washed
'Peter has washed the car.'				
a. Er	hat	**den**	**Außenspiegel**	ausgelassen.
He.NOM	has	the.ACC	side mirror	left-out
'He left out the side mirror.'				

b. **Den**	**Außenspiegel**	hat	er	ausgelassen.
The.ACC	side mirror	has	he.NOM	left-out

'The side mirror, he left out.' (Weskott et al., 2011: 7)

In (1.19) the expressions *den Wagen* ('car') and *den Außenspiegel* ('side mirror') are in a whole-part relationship belonging to a common poset. The object noun phrase *den Außenspiegel* in (1.19b) is at the beginning of the sentence, whereas in (1.19a) it is in its 'canonical' position. Weskott et al. (2011) show in their perception experiments that the sentence in (1.19b) is highly preferred in the context of (1.19C). Thus, the sentence-initial position for the object noun phrase is licensed by the poset "car and its subparts." In terms of this study, mentioning *the car* activates *the side mirror* in the discourse and this semi-active state of the referent *side mirror* might be reflected in the sentence-initial position of the referring expression.

1.3.3. Accentuation and free word order

A number of studies (Vallduví & Engdahl, 1996; Lambrecht, 1994; Ladd, 2008, Van Valin, 1999) observe that some languages require a fixed position for the nuclear pitch accent somewhere in the intonational phrase. Enric Vallduví and Elisabet Engdahl (1996) propose that in Catalan, prosodic prominence can occur only at the end of an utterance. In addition, they suggest that in Catalan, word order needs to adjust to the strict structure of the intonational phrase in order to highlight focus by prosodic prominence. For this adjustment a sentence constituent in focus is shifted to the sentence-final position in order to carry an accent. Lambrecht (1994: 318) suggests the same for Italian as shown in example (1.20) (capital highlights the position of nuclear accent, bold marks the subject noun phrase that is switching its position).

(1.20)

A. What's the matter?

a. My NECK hurts.

b. Mi fa male il **COLLO**.

B. How's your neck doing?

a. My neck HURTS.

b. Il **collo** mi fa MALE.

Observe in (1.20) that in context B English deaccents the subject noun phrase *neck* and accents the verb *hurts*. In Italian, however the word *collo* switches the sentence position from the end of the sentence to the beginning of the sentence. The proposal is that in Italian the constituent in focus requires pitch accent that is obligatorily located at the end of an intonational phrase. The phenomenon demonstrated in (1.20) is captured by a distinction between *plastic* and *non-plastic* languages (Vallduví & Engdahl, 1996). In this terminology English represents a plastic language because it enables accent shift (see section 1.2.1 for explanation) for focus highlighting. To the contrary, Catalan or Italian that do not appear to have this possibility are non-plastic languages (Vallduví & Engdahl, 1996; Ladd, 2001). However, free word order does not necessarily imply that a language has a fixed position for sentence accent. Robert D. Van Valin (1999) reports free word order languages that do enable accent shift, such as Russian. In contrast to Lambrecht's suggestion shown in example (1.20), Timothy Face and Mariapaola D'Imperio (2005) argue that Italian, being a free word order language, successfully implements pitch accents and accent shift in order to convey focus.

Hungarian belongs to the languages that phonologically have free accent placement, however, the location of nuclear pitch accent is determined by structural focus in Hungarian (Siptár & Terköczy, 2000; Varga, 2002; Szendröi, 2003). Thus, nuclear accent always appears before the verb, see example (1.21).

(1.21)

C1: Katalyn knows Kenzi.
A1. **IMRE** ismeri Kenzit.
Imre knows Kenzi.ACC
A2. Kenzit **IMRE** ismeri.
Kenzi.ACC Imre knows
'No, IMRE knows Kenzi.'

C2: Imre knows Katalyn.
B1. **KENZIT** ismeri Imre.
B2. Imre **KENZIT** ismeri.
'No, Imre knows KENZI.

(Siptár & Terközcy, 2000: 47)

(C1) in (1.21) focuses on subject noun phrase *Imre,* whereas (C2) on the object noun phrase *Kenzi*. Both subject and object noun phrase can occur in the sentence-initial and sentence-medial position, but they have to precede the verb *ismeri* ('knows'). They are appropriate for the given contexts (C1, C2) only in preverbal position. Observe that nuclear pitch accent also occurs before the verb and shifts from the initial position to the medial one in (A2) and (B2). Varga (2002) agrees, but he speaks rather about deaccentuation of the verb than about the nuclear accent in the preverbal position.

As already mentioned, Välimaa-Blum (1988, 1993) and Vilkuna (1998) account for different word orders in Finnish as being grammatical constructions with particular semantic and pragmatic value. In addition, Välimaa-Blum (1993: 125) finds that the accentuation pattern is a formal concomitant of each word order construction. In her studies, Välimaa-Blum (1988, 1993) shows for Finnish that the verb-final and verb-initial word orders are by default produced with phrase-initial nuclear pitch accent. Recall from section 1.3.2 that these word orders have focus either on the beginning of a sentence (verb-final word orders) or 'emphasize' a whole proposition (verb-initial word orders). However, Välimaa-Blum (1993) makes a relevant insight:

> *"a construction can be used to express other meanings or functions, too, depending on the context. And if so, the default reading must be cancelled and this can be done by morpholexical means and/or by intonation."* (Välimaa-Blum, 1993: 125)

Thus, if a construction is embedded into a context with pragmatic implications different from the construction, then it can be accommodated with those implications either by morpholexical means or by appropriate placement of a sentence accent. As shown in example (1.22), Välimaa-Blum (1988) provides a vivid demonstration for overriding the implications provided by the construction.

(1.22)

a. SORSIA lammessa ui.
   duck.PART.PL pond.INE.SG swam
   'DUCKS swam in the pond.'

b. Sorsia lammessa ui **eilen**.
   duck.PART.PL pond.INE.SG swam yesterday
   'Ducks swam in the pond yesterday.' (Välimaa-Blum, 1988: 77)

Both word orders in (1.22) have the verb *ui* (‘swam’) in the third position. Example (1.22a) is a verb-final order that has the expression *sorsia* (‘ducks’) in focus. As soon as an additional adverb, *eilen* (‘yesterday’), is added to the construction, as in (1.22b), focus is shifted from sentence-initial constituent to sentence-final constituent. With regard to this, the question arises whether the accent shift in (1.22a) would cause an analogous change in the focus of an utterance. Välimaa-Blum (1988, 1993) does not provide data for that, but Stavros Skopeteas, Caroline Féry and Rusudan Asatiani (2009) have investigated similar question in Georgian – an unrelated language that also has an extremely free word order.

Georgian has quite a strong preference to have focus located immediately before or after the verb. In addition, verb-initial word orders strongly cue the verb in focus. Skopeteas et al., (2009) investigated whether a nuclear pitch accent either on a constituent that is not verb-adjacent or on a constituent that is not sentence-initial verb would cue focus on that constituent. Thus, the question was which grammatical means – word order or pitch accent – interacts stronger with information structure in Georgian. Their results showed that pitch accent on a constituent that was not verb-adjacent cued efficiently focus on that constituent. This suggests that the pragmatic implication specific to a verb-adjacent constituent can be overridden by sentence prosody, compatibly to a proposal by Välimaa-Blum (1993). However, in verb-initial word orders it was impossible to cue focus on a non-verbal constituent. The data from Georgian imply, thus, that some word orders (constructions) might have stronger pragmatic implications than others.

1.3.4. Interim summary

To conclude, free word order means that all the constituents involved occur in all their (or in a few) logically possible permutations (Välimaa-Blum, 1988: 61; Siewierska, 1998), but it does not necessarily mean that the occurrence of these orders is ungoverned. In section 1.3.2 it was discussed that the information-structural categories involved in constituent ordering can be focus, identifiability and activeness. The effect of focus was discussed based on Hungarian, the effect of identifiability based on Finnish and Russian and the effect of activeness based on the data of English and

German word order. Section 1.3.3 observed the interaction between information structure, free word order and the placement of nuclear accent. To us, Hungarian data (Siptár & Terköczy, 2000; Varga, 2002) and Finnish data (Välimaa-Blum, 1988; 1993) serve as two possible theoretical models for interaction between accent and free word order. In Hungarian, nuclear pitch accent shifts together with a referring expression in focus (Siptár & Terköczy, 2000; Varga, 2002). In Finnish, word order might have a specific accentuation pattern and pragmatic implications but they can be overridden by pragmatic implications of a linguistic context (Välimaa-Blum, 1993). In the following chapters, one of the aims is to answer the question whether pitch accent can be used as a cue for overriding a pragmatic implication embedded in a particular word order.

1.4. Object of the study: Estonian

Estonian belongs to Finno-Ugric languages (Abondolo, 1998; Erelt, 2003) and about 922,000 speakers in Estonia and about 160,000 speakers outside Estonia speak it as a native language (Estonica.org, 22th September 2015). Estonian belongs to the Finnic branch of the Finno-Ugric languages (Viitso, 2003) and it is closely related to Finnish and less closely to Hungarian. Despite the small number of speakers and the small area of the country (about 45,277 km^2), a number of dialects can be detected (up to 120) that belong to three main dialect groups: North-Eastern Coastal Estonian, North Estonian and South Estonian (Pajusalu, 2003). Nowadays the majority of Estonians speak Standard Estonian that was developed in the beginning of the 20th century on the basis of Northern-Estonian dialects in order to serve as a norm for written Estonian (Laanekask & Erelt, 2003).

Estonian has gained a lot of attention as a *quantity language* (Ariste, 1939; Lehiste, 2002), in which there is a short-long vowel and consonant contrast at the lexical level, like in Finnish and Estonian (Lehiste, 1965; Lehiste, 1960). The research on Estonian prosody has concentrated a lot on quantity variation, whereas sentence intonation is less investigated (Asu, 2004).

Estonian is typologically an agglutinating language with strong flectional tendencies (Erelt, 2003). For nominal constituents it has 14 morphological suffixes: three for grammatical cases and 11 for adverb cases that cover relations in time and space like adpositions in many other languages (Viitso, 2003: 32). Among the

grammatical cases, nominative (with null-sufix) is the case of the subject noun phrase and genitive or partitive the case of the object noun phrase (Viitso, 2003: 32). Like typological data predicts (Siewierska, 1998), Estonian is a *free word order* language in which all the logical permutations of subject noun phrase, object noun phrase and verb are grammatical (Tael, 1988; Erelt et. al., 1993; Lindström, 2006). The following sections 1.4.1 and 1.4.2 give a short overview of Estonian quantity and sentence intonation. Section 1.4.3 discusses Estonian word order variants in connection to information structure.

1.4.1. Estonian quantity

Estonian sound inventory consists of nine vowel phonemes /i, y, e, ø, æ, u, o, ɤ, a/ and 17 consonant phonemes: /p, t, t^j, k, m, n, n^j, r, f, v, s, s^j, ʃ, h, l, l^j ,j/ (Asu and Teras, 2009). All vowel and consonant phonemes can occur as short and long (single *vs.* double symbols, e.g. /e/ *vs.* /ee/) and the vowel phonemes combine into 36 diphthongs that can only be long (Asu & Teras, 2009). Word stress is fixed on the first syllable (Lehiste, 1960; Eek & Meister, 2004), with few exceptions, like *aitäh* (‘thank you’).

The segmental constituency of stressed and unstressed syllables is phonotactically highly constrained. Short, long and overlong vowels can occur in an initial stressed syllable rhyme, whereas only short vowels can occur in subsequent unstressed syllables (Viitso, 2003: 25). All nine vowels can occur in stressed syllables or as the first component of a diphthong, but only five vowels /a, e, i, o, u/ are allowed in non-first unstressed syllable or as a second component of a diphthong (Asu & Teras, 2009). Secondary stressed syllables are allowed to consist of long vowels but only three types of diphthongs: /ɑi/, /ei/ and /ui/ (Lehiste, 1997; Asu & Teras, 2009). Word-internal consonant clusters (between the vowels) can consist of two to five consonants (Viitso, 2003: 23); the last consonant of the cluster or a single consonant belongs to a syllable onset of the unstressed non-first syllable (Lehiste, 1997). Thus, segmental constituency of stressed syllables is more variable due to long vowels, diphthongs and complex codas than unstressed syllables. This phonologic fact could probably explain the duration as a main acoustic correlate for word stress in Lippus et al. (2014).

In terms of rhythm, Estonian has been classified as a *stress-timed language* (Eek & Help, 1987; Asu & Nolan, 2006) in which stress beats appear in approximately

regular intervals (see Fletcher, 2013 for discussion of rhythmical types of languages). Arvo Eek claims that there is a tendency in Estonian to regulate the duration of segments within a (stress) foot (Eek, 1990: 256) that consists of one, two or three syllables (Eek, 1990: 252). *Estonian quantity* is a combination of duration, intensity and pitch (Lehiste, 1997) and as such constitutes the main *prosodic property* of the foot. Thus, phonemes in Estonian can occur as short and long, but at the level of the foot there is a three-way distinction of short (Q1), long (Q2) and overlong (Q3), see Table 1.1 for illustration. Observe in Table 1.1 that in case of vowels, orthography does not distinguish between Q2 and Q3.

Table 1.1 Examples of Estonian word triplets with vocalic and consonantal quantity

		Vowel-quantity	**Consonant-quantity**
Q1	IPA	[vɑ.lu]	[vɑ.kɑ]
	Orthography	valu ‘pain’ NSg	vaga ‘pious’ NSg
Q2	IPA	[vɑɑ.lu]	[vɑk.kɑ]
	Orthography	vaalu ‘haystack’ GSg	vaka ‘granary bin’ GSg
Q3	IPA	[vɑɑ:.lu]	[vɑk:.kɑ]
	Orthography	vaalu ‘whale’ PPl.	vakka ‘granary bin’ PSg

The monosyllabic foot is assigned to be in Q3 with an exception of the function words and short forms of the pronouns. It is mainly a theoretical procedure, because monosyllabic words do not show consistent phonetic characteristics (Eek & Meister, 2003) and they tend to combine into a di- or tri-syllabic foot with a preceding foot (Eek, 1990). Thus, the three-way quantity distinction occurs in bi- or tri-syllabic feet in which the relevant cues of duration and pitch appear during the first two syllables (Lehiste, 1960; Eek, 1990; Lippus et al., 2013). A foot containing a short consonant (C) or vowel (V) is in Q1 (/vɑlu/ and /vɑkɑ/ in Table 1.1). A foot containing a long vowel, diphthong or a long consonant (/vɑɑlu/ and /vɑkkɑ/ in Table 1.1) can be either in Q2 or Q3 ([vɑɑlu] *vs.* [vɑɑ:lu] and [vɑkkɑ] *vs.* [vɑk:kɑ]). The distinction between Q2 and Q3 emerges in different duration ratios of the syllables and in different placement of the pitch peak in relation to the first stressed vowel (Lehiste, 1960; Eek, 1974; Mihkla & Kalvik, 2011; Lippus, et al., 2013).

The duration of the second unstressed syllable is inversely proportional to the duration of the first stressed syllable: the longer the first syllable, the shorter the second

one (Eek, 1990: 261). The phenomenon is best described by constant duration ratios of syllables: 2:3 for Q1, 3:2 for Q2 and 2:1 for Q3 (Lehiste, 1960; 1997). The duration ratios reflect the tendency of the overall duration of the disyllabic foot with different quantities to be relatively alike, Q1 foot being slightly shorter. The pitch peak in Q1 and Q2 is aligned with the second half of the first syllable, whereas in Q3 it is aligned with the first half of the first syllable (Lehiste, 1960; Eek, 1974; Mihkla & Kalvik, 2011; Lippus et al., 2013). In terms of peak alignment, there is a late peak in Q1 and Q2 foot but an early peak in Q3 foot.

The main reason why it is emphasized in literature that the three degrees of Estonian quantity should be accounted for in the domain of foot is the experimental evidence that native Estonians are unable to distinguish between the Q2 and Q3 without a second unstressed syllable (Eek & Meister, 1997). In addition, it has been shown that the tonal cue is crucial for quantity identification in various perception tasks (Lehiste, 1997; Eek, 1980; 1983; Lippus, 2007; 2009; 2011; Salveste, 2010). Therefore, it can be concluded that quantity in Estonian is a prosodic property of a disyllabic foot that is cued by the syllable ratio of stressed and unstressed syllables and by early *vs.* late peak alignment.

1.4.2. Estonian intonation

According to Eva-Liina Asu (2004; 2005), Estonian distinguishes between six types of different pitch accents, two of which are monotonal (H*, L*) and four bitonal (H*L, HL*, L*H, H!H*). Three pitch accents are restricted to occur only in a nuclear position (H*, L*H, L*); the other three can appear in prenuclear as well as in nuclear position (H*L, HL*, H!H*). With respect to pitch cue of Estonian quantity, the issue of tonal alignment in pitch accents is of particular interest; the question arises how these interact with quantity dependent peak alignment. For L*H pitch accent in which the intonation in the stressed vowel is low, Eva Liina Asu and Francis Nolan (1999) have found that the quantity dependent peak alignment is not realized. The low target occurs right at the end of the vowel and the peak is located in the beginning of the next syllable. Asu and Nolan (2007) have found that the most frequent pitch accent is high falling accent (H*L). This might be related to the logical possibility that the H*L pitch accent fits with quantity-dependent tonal cue.

Since the low boundary tone is most frequent, the boundary tone can be left unspecified (%), except for the nuclear rise where L* is followed by a high boundary tone H% (Asu 2004). Therefore, the Estonian intonational phonology comprises of low and high boundary tones to mark boundaries of intonational phrases. There is no division into intermediate phrases and there are no special tunes for phrase accents.

Statements and questions can both be produced with H*L pitch accent with a low boundary tone (Asu, 2004: 56). However, there is some evidence that the questions might be signalled by an expanded pitch span in Estonian (see Vende, 1975; 1982). The existence of a rising intonation has been rejected for Estonian for a long time (Asu, 2006; Keevallik, 2003). In reality, high or rising boundary tones appear to carry relevant pragmatic function of continuation in the interaction of speech participants (Asu, 2006). In spontaneous dialogues, rising accents were shown to occur most frequently with feedback particles such as *jaa* ('yes', 'right') or *mhmh* (Asu, 2006). Kasterpalu (2013) provides an insight about different intonational tunes on *jaa-jaa* ('yes-yes', 'right') as a feedback particle in sales negotiations. The rising tone on the *jaa-jaa* particle signals that the information provided by the speech partner was old information; the falling tone signals that the information provided by the speech partner was new information to the speaker.

The prosodic means of focus in Estonian have gained only recent interest. Heete Sahkai, Mari-Liis Kalvik, Meelis Mihkla (2013b) have investigated whether different information-structural categories such as focus or topic are signaled with different types of pitch accent. They found that focus and topic are signaled with pitch prominence (several types of pitch accents were attested) but there was no correspondence between pitch accent type and information-structural category. Sahkai et al. (2013a) found that focus is signalled with clear prosodic prominence (longer segment durations) but not necessarily by higher F0 or F0 expansion.

1.4.3. Free word order

Estonian is reported to have free word order (Vilkuna, 1998; Remmel, 1963) and most of the current researchers of Estonian syntax agree on it (see Ehala, 2006; Lindström, 2006, Erelt et al., 1993 for example). As discussed in chapter 1.3, one of the reasons for rigid word order might be the need to encode thematic relations in a sentence, as it is the

case in English. Estonian encodes subject and object noun phrase with case marking: nominative (with null-suffix) is the case for the subject noun phrase and genitive or partitive is the case for the object noun phrase (Viitso, 2003: 32), see example in (1.23).

(1.23)

Mees-Ø	armasta-b	nais-t
man-Sg.Nom	love-Sg3	woman-Part.Sg

'The/A man loves the/a woman.'

Observe in (1.23) that the object noun phrase *naist* ('woman') is marked with a suffix *-t,* whereas the subject noun phrase is not marked. Thus, the rich morphological system appears to enable free placement of words in a sentence. The main force for the order of main sentence constituents (subject or object noun phrase) in Estonian is information structure (Tael, 1988; Erelt et al., 1993; Lindström, 2002; 2004; 2006).

Erelt et al. (1993: 13–14) list three structural mechanisms governing Estonian word order: *theme-rheme* structure (not mentioned explicitly, but probably in the sense of Halliday, 1967a), *definiteness* and *focus*. Theme-rheme structure is a logical structure that gives a sentence a feeling of completeness (Erelt et al., 1993: 13). Theme is a portion of a sentence about which the rheme makes an assertion (Erelt et al., 1993: 13); see illustration in (1.24) for an effect of theme-rheme structure in Estonian word order.

(1.24)

Theme	**Rheme**		
a. **Peeter**	luges	"Sõrmuste	isandat".
Peter	read.SG3.PST.	ring.PL.GEN	lord.PRT
'Peter read the "Lord of the rings."			
b. "Sõrmuste	isandat"	luges	**Peeter**.
ring.PL.GEN	lord.PRT	read.SG3.PST	Peter
'Peter read the "Lord of the rings."			

Observe in (1.24) that in Estonian the sentence *Peter read the "Lord of the rings"* can be rendered with two word orders (SVO and OVS) depending on what the speaker wants to make as a 'point of departure' (theme) of his utterance.

Estonian does not have any articles for definite and indefinite noun phrases, but the definiteness can be optionally encoded in word order according to

Erelt et al. (1993). A definite noun phrase precedes an indefinite noun phrase in a sentence and often it means that the noun phrase in the beginning of a sentence is definite (Erelt et al., 1993: 13).

Erelt et al., (1993) observe impressionistically that a constituent in focus is signalled with sentence stress, see example (1.25).

(1.25)

Sinu	kasvatasin	ju	MINA	üles.
you.Part	raise.Sg1.Pst	mod.particle	I	up

'I raised you.'

(Erelt et al., 1993: 14)

In example (1.25), capitals mark the position of the sentence stress. It is to be noticed that in (1.25) the subject *mina* ('I') in addition to carrying sentence stress is also in a non-canonical position (after the verb *kasvatasin* 'raised'). As implied by the example in Erelt et al. (1993), the prosodic prominence appears to shift together with word order.

For further discussion of Estonian word order, main sentence types need to be introduced (Erelt et al., 1993 refer to clause type) because the information-structural effect of a particular sentence position depends on the type of a sentence. The three sentence types are defined on the basis of how the grammatical function of the subject interacts with the semantic agent[4] and with theme-rheme structure (Erelt et al., 1993: 14). The examples of three sentence types are given in (1.26).

(1.26)

a. Jaan	kirjutab	raamatut.
John	write.SG3	book.SG.PART
'John writes a book.'		
b. Jaanile	meeldib	tantsida.
John.ALL	like.SG3	to dance
'John likes to dance.'		
c. Peenral	kasvab	lilli.
flowerbed.AD	grow.SG3	flower.Pl.PART
'The flowers grow in the flowerbed/ garden'		

[4] *The semantic agent* refers participant role of actor or initiator of the event expressed by the sentence (see Van Valin, 1993; 2009).

Example (a) in (1.26) is called an *unmarked basic* clause (Erelt & Metslang, 2006: 254; Erelt et al., 1993: 14). In sentence (a) the grammatical subject *Jaan* is a referent that initiates the action (agent) and the 'point of departure' of an utterance (theme). Thus, in an unmarked basic sentence the theme and the agent coincide with the grammatical subject. Sentence (b) in (1.26) is called an *experiential* clause (Erelt & Metslang, 2006: 255; Erelt et al., 1993: 14). It can be observed in (1.26b) that the sentence-initial constituent *Jaan* is not the grammatical subject, because it is marked with one of the cases for adverbs – adessive that prototypically marks location. In example (1.26b), the grammatical subject is not located in the beginning of the sentence and can be even left out from the experiential sentence (Erelt et al., 1993). However, semantically the sentence has an actor *Jaanile* ('John') that functions as the theme of the sentence. Thus, in the experiential sentence, the grammatical subject coincides neither with the theme nor the semantic agent. However, the theme and the agent coincide with each other. Sentence (c) in (1.26) is called an *existential* clause (Erelt and Metslang, 2006: 255; Erelt et al., 1993: 14). Similarly to the experiential sentence, the existential sentence does not contain any constituents in nominative case that can function as the grammatical subject. In the existential sentence, the semantic agent and the theme do not correspond to each other as they did in the experiential sentence. The constituent *lilli* ('flowers') is the agent (the only animous referent that can 'do' the growing) in (1.26c), whereas *peenral* ('flowerbed') is the theme (Erelt et al., 1993: 14). It should be noted that the semantic agent is in the position of rheme.

In brief, the theme ('point of departure') is located in the beginning of the sentence, whereas the rheme is at the end of the sentence. A special sentence structure – the *existential sentence* – can be used to make the semantic agent to the rheme of the sentence. To achieve the same effect for the subject noun phrase/semantic agent in the transitive sentence (1.24a), the subject is shifted to the end of the sentence as it occurs in (1.24b). This is often referred to as *inversion of subject noun phrase*.

The variation of word order in Erelt et al. (1993) is described from the perspective of the theme-rheme structure. Recall that for Halliday (1967a) the theme-rheme structure is not related to the information structure or to the information value of the sentence components. Yet Gundel (1999) finds, using different terminology though, that the rheme introduces new referents to the discourse. The theoretical account of Estonian word order (Erelt et al., 1993) appears to involve the informational value of

the sentence components as well. The possibility to place the sentence constituents to the end of the sentence is, thus, connected to signalling information that is due to newness also in focus.

The theory presented above is congruent with a corpus study of written Estonian carried out by Kaja Tael (1988). On the basis of her corpus analysis, Tael (1988) describes two main information-structural effects on word order. Firstly, the topic (probably in terms of Prague School (see Sgall et al., 1986) or Gundel (1985) but not stated explicitly) is located in the beginning of the sentence. If the subject noun phrase does not refer to the referent suitable for the topic, the sentence occurs to have an inversion of the subject noun phrase. Secondly, the sentence-initial position can also be 'emphatic', called as *emphatic topic* in Tael (1988: 11, 38). Tael (1988: 11) claims that the sentence-initial constituent is emphatic in majority of sentences that have an inverted subject in them (X_nVS orders). However, the emphatic interpretation occurs to depend on the sentence type. According to data in Tael (1988: 10–11), the emphatic reading of a sentence-initial constituent is possible in transitive and experiential sentence types but not in an existential sentence (see 1.26c). As was seen above, the existential sentence is a special construction that enables to signal the information contained in the semantic agent as rheme of the sentence or as new to the discourse (Erelt et al., 1993).

Tael (1988) does not distinguish between the grammatical subject and the semantic agent in her analysis as it is done in Erelt et al., (1993) but she shows that a sentence with an inverted subject is the most frequent in experiential and existential sentence types (pp. 13). In her account (Tael, 1988), this is evidence that the sentence-final position in the experiential and the existential sentence is neutral in terms of information-structure. She goes further, stating that word order with an inverted subject is only grammatical word order for the existential sentence type. Interestingly, Välimaa-Blum (1988) finds for Finnish that the existential sentence has free word order (see chapter 1.3.2 for more details).

In addition, Tael (1988: 37) finds that the referent of the subject noun phrase that follows the verb is new to the discourse and unidentifiable to an addressee. In a corpus study carried out on spontaneous speech, Liina Lindström (2002: 99) finds compatibly with Tael (1988) that the subject noun phrase referring to the referent that is new and unpredictable for a discourse occurs frequently in sentence-final position.

1.4.4. Interim summary

This subsection presented the main characteristics of Estonian prosody. The understanding of the prosodic manifestation of Estonian quantity is important for investigation of Estonian sentence intonation. The only in-depth study of Estonian intonational phonology (Asu 2004) comprises of six pitch accents and two boundary tones. The most frequent pitch accent has shown to be H*L (Asu & Nolan, 2007). Estonian is a language with highly flexible word order in which all the permutations of the sentence constituents are grammatical (Lindström, 2004; 2006). The pragmatic factors of discourse determine the appropriate usage of word order. If the speaker wants to present information as new to the discourse (focus), then he can place a sentence constituent to the end of the sentence (Tael, 1988; Erelt et al., 1993; Lindström, 2006). Also sentence-initial position can highlight focus in some sentence types (see Tael, 1988). In addition to focus, a sentence-final position can also signal that a referent is unidentifiable to a listener.

The sentence accent appears to be a less investigated part of Estonian information structure. Sahkai et al., (2013ab) provide some tentative data that pitch prominence is also an optional cue for focus. If a language has several linguistic devices for focus marking as Estonian does with both pitch prominence and word order, the question arises, how these devices interact with each other? Is the position of nuclear pitch accent shifted together with some sentence constituents (like in Hungarian) or does it have its own independent interface with information structure like implied by the hypothesis of overriding in Välimaa-Blum (1988; 1993)? The experimental studies presented in chapters 2, 3, 4 and 5 tap into these questions. The view that pitch accent is an important cue for focus and that deaccentuation signals givenness is going to be defended. In doing so, it is predicted that the pragmatic functions of word order can be overridden by the placement of nuclear pitch accent, like suggested by Välimaa-Blum (1993).

1.5. Conclusion

The sections of this chapter have discussed information structure, intonation and word order. In section 1.1 the theory of information-structural categories was presented. Lambrecht (1994) provides two important recognitions for the theory of information structure. First, the linguistic form that the discourse function might take is not the

function itself. In other words, pitch accent as a grammatical means is not equal to pragmatic function focus. Second, the simple dichotomy given-new cannot account for different information-structural values that the speech participants appear to acknowledge by their usage of linguistic means. This means that the theory of information structure needs to consist of more categories and subcategories. A number of other studies (Prince, 1981; 1992; Baumann, 2006; Baumann & Grice, 2006; Baumann & Riester, 2012) support this view. Lambrecht's (1994) four-way account outlines one set of possible major categories.

Focus is the information that the speaker wants to present as new to the hearer (Halliday, 1967a; Lambrecht, 1994). Activeness and identifiability refer to properties of information status that referents have in the minds of speakers (Chafe, 1974; 1976; 1987; Lambrecht, 1994). In linguistic research, it is not possible to estimate 'objectively' the consciousness of the speech participants but it might be possible to investigate and estimate the factors that influence the speaker's choice of linguistic devices in a particular discourse.

Section 1.2 discussed sentence accent. Pitch prominence was attested to be the main cue of focus and the deaccentuation of activated or given information in intonation languages (e.g. English and German). Section 1.3 discussed word order in connection to the information structure. Three information-structural factors were recognized to affect word order: focus in the example of Hungarian (É. Kiss, 1995), identifiability in the example of Finnish (Välimaa-Blum, 1988) and activeness in the examples of English (Ward and Birner, 2004) and German (Weskott et al., 2011). Section 1.4 described Estonian prosodic structure and the theory of word order in order to provide preliminary knowledge for understanding the experiments reported in chapters 2, 3, 4 and 5.

Finally, for Lambrecht (1994), information structure, phonology (also intonational phonology) and syntax (in terms of word order) are three components of grammatical sentences that *"are seen not as hierarchically organized independent subsystems but as interdependent forces competing with each other"* (Lambrecht, 1994: 12). In other words, different grammatical components are given different weight in a language. The idea of competition underlies the experiments that were carried out in order to establish whether information structure or word order interacts more strongly with sentence intonation.

2. Focus perception in Estonian: syntactic or prosodic?[5]

Abstract

It is known for *intonation languages* that the placement of nuclear pitch accent highlights the location of sentence focus (Cooper et al., 1985; Eady & Cooper, 1986; Pierrehumbert & Hirschberg 1990, Breen et al, 2010 for English; Baumann, 2006; Féry & Kügler, 2008 for German; and Swerts et al. 2002 for Dutch). In free word order languages the displacement of the word in focus and morphological markers are used to highlight focus (Vallduví, 1993; É. Kiss, 1995; Rizzi, 1997; Féry & Krifka, 2008). In Estonian, both prosodic prominence and position in a sentence are reported to cue focus (Erelt et al., 1993). The aim of the study was to investigate whether pitch prominence elicits perception of focus in Estonian and, if so whether sentence-final position or pitch accent is a stronger cue for focus.

A perception test with a forced choice task was run. Native Estonian listeners were presented with short narrative excerpts that ended with an indirect question with a narrow focus either on an object or an adverb. After reading an excerpt, listeners performed a congruence-matching task in which they were asked to decide which of the two recordings of the same sentence was semantically the most congruent with the written excerpt presented on the computer screen. The stimuli were constructed in a way that the word (object or adverb) in focus was either sentence-final or sentence-medial and carried either a pitch accent or was unaccented. The stimuli were combined into pairs and the listeners had to choose between two sentences in which, for example, the word in focus was either sentence-final accented or sentence-medial accented.

The results showed that a sentence-final position alone could not cue sentence focus. The listeners most frequently matched the stimulus-sentence in which the word carried a pitch accent with focus implied by the context. The main outcome is that L1-Estonian listeners attend to intonational prominence and make effective use of the placement of nuclear pitch accent in identifying focus in a sentence. The study shows that intonational prominence is an optional cue for focus in a language that marks information structure with word order.

[5] A version of this chapter was published in the *Proceedings of the XIth Conference of Nordic Prosody* (Salveste, 2013).

2.1. Introduction

English, German and Dutch are reported to belong to intonation languages that tend to mark information structure (focus *vs.* non-focus) of the constituent tonally (Brown, 1983; Cooper et al., 1985; Eady & Cooper, 1986; Pierrehumbert & Hirschberg 1990, Breen et al, 2010 for English, Kohler 1991, Baumann, 2006; Féry & Kügler, 2008 for German & Swerts et al. 2002 for Dutch). In other types of languages, the information status is marked by a change of the position of a word in a sentence (Rizzi, 1997; Kiss, 1995). The study investigates for Estonian whether native speakers interpret either pitch accent or sentence-final position as a focus marker.

In Hungarian, the constituent in focus occurs before the verb (É. Kiss, 1995; Mycock, 2010, Szendröi, 2003). The same pattern appears in Georgian (Skopeteas et al., 2009). In Greek, the word in focus needs to occur in the beginning of the whole sentence (Keller & Alexopoulou, 2001). In Finnish, the 'emphatic' focus is encoded in the beginning of the sentence; the constituent in the new information focus is placed to the end of the sentence (Vilkuna, 1995; 1998; Välimaa-Blum, 1988). Typological overviews (Vallduví & Engdahl, 1996; Van Valin, 1999) propose that there are languages that use intonation only, languages that use syntax only and languages that use both of them. However, it is rare that a language fits in the boundaries of these categories completely. Even English or German, claimed to be intonation languages, use word order variation for information-structural purposes (Birch & Ward, 2004; Fanselow & Lenertova, 2011; Weskott et al., 2011). Even if a language is claimed to use solely syntactic means, the word in focus is still presumed to carry an accent (Siptár & Terköczy, 2000; Varga, 2002; Szendröi, 2003). In case of Italian or Hungarian, researchers even argue (Vallduví, 1992; Vallduví & Engdahl, 1996; Zubizaretta, 1998; Szendröi, 2003; Samek-Lodovici, 2005) that the metrical structure (the abstract strongest prominence has fixed position in a phrase) determines the need to reorder the constituents.

Martti Vainio and Juhani Järvikivi (2006) investigated categorical effects of gradually manipulated F0 contour in perception of Finnish sentence prominence. In the first experiment, they found for a sentence consisting a verb (V), a manner adverb (M) and a place adverb (P) in the order of VMP that the number of accents on the place adverb the listeners reported was equal to the number of accents on the manner adverb.

The perceived accent correlated to the position of the nuclear accent. However, when the word order was changed from VMP to VPM in the second experiment, the participants reported significantly more accents on the manner adverb than on the place adverb, although the tonal representation of sentences was kept the same. Thus, the study showed that in general Finnish L1-listeners follow the location of the nuclear pitch accent in detecting the focus of a sentence, but the sentence-final position can cause perception of prominence for some type of constituents irrespectively of the location of the nuclear pitch accent. This suggests that the structural properties such as word order can induce prominence perception without any phonetic cue.

In this study the interaction of different linguistic means of focus in another free word order language – Estonian – is investigated. Estonian is typologically related to Hungarian and Finnish (Abondolo, 1998). Similarly to Hungarian and Finnish, morphological cases encode syntactic relation and Estonian is also reported to use word order for the transmission of information structure. By some researchers, Estonian has also been considered to be a discourse-configurational language or to share properties with discourse-configurational languages (Vilkuna, 1998; Ehala, 2006). However, Estonian does not seem to have an obligatory focus position like Hungarian does. Sentence constituents tend to occur in some positions of a sentence, if they have a certain information-structural status: for focus, the sentence constituent can be placed either to the beginning of a sentence or to the end of a sentence (Tael, 1988; Erelt et al., 1993; Lindström, 2004; 2006). Corpus studies of spoken Estonian by Lindström (2002) have shown that the position of subject noun phrase in a sentence depends partly on whether the subject refers to new information. In Lindström (2002), subject referred to a known/mentioned referent in 81% of SV word-orderings, and to a new referent in 80% of VS word-orderings.

Sparse empirical data exists for prosodic marking of information status in Estonian. There are a few studies exploring the acoustics of accent (Sahkai et al., 2013ab), not to mention the prosodic highlighting of focus in connection with different word orders. There is some evidence that pitch accent occurs preferably either at the left or at the right edge of the intonational phrase. Eek (1983: 483) impressionistically notes that the tonal properties of Estonian quantities only occur at the beginning or at the end of a phrase, whereas in the phrase-medial position they tend to neutralize phonetically. A recent study by Asu et al. (2009) confirms this observation. Thus, the theoretical

sentence-initial position of non-focus/contrastive focus and the sentence-final position of focus are frequently accompanied by intonational prominence in Estonian. The both of the means – syntactic or prosodic – appear to be available in the language, but which of them constrains the context of an utterance stronger?

An experiment was designed to test whether sentence-final position or pitch accent elicits focus perception. The following subsection investigates the strength of the pitch accent against the sentence-final position in verb-adverbial-object (VAO) and verb-object-adverbial (VOA) sentences. The predicted focus cue is the location of the nuclear pitch accent.

2.2. Experiment

A perception experiment with a forced choice task was run. L1-Estonian listeners were presented short narrative excerpts that ended with an indirect question with a narrow focus either on the object noun phrase or on the adverb. After reading the excerpt, they were asked to decide which of the two recordings of the same sentence was semantically the most congruent with the written excerpt presented on computer screen. In the following, the recordings will be referred to as *stimulus*, the excerpts as *context*.

2.2.1. Stimuli

12 sentences consisting of a verb (V), an adverb (A) and an object (O) (VAO) were constructed and used as stimuli (see example 2.1). Another 12 sentences were derived from them by swapping the order of A and O: VOA; see the example (2.2). The prosodic structure of stimuli was kept alike. They were composed of words consisting of voiced sounds or sounds that tend to become voiced between voiced segments. The V at the beginning of the sentence was tri-syllabic word of long quantity (Q2). The disyllabic locative A was Q2 word. O was also disyllabic, but of overlong quantity (Q3).

(2.1)

Värvi-me	õue-l	laeva.
to paint-1PL	garden-AD	boat.PART

'Let's[6] paint the boat in the garden.'

(2.2)

Värvime	laeva	õuel.
to paint.1PL	boat.PART	garden.IN

'Let's paint the boat in the garden.'

Sentences with different word order (VAO and VOA) as read by a trained phonetician (the author) were recorded and manipulated in such way that the nuclear pitch accent occurred either sentence-medially or sentence-finally. The F0 contour of the recorded sentences was manipulated in Praat (Boersma & Weenink, 2012).

Only one very high pitch peak occurred per sentence. The placement of the peak took into account the fact that the peak alignment in Estonian varies in relation to quantity (Lehiste, 1960; Eek, 1983; Lippus et al., 2013). In Q3 words, the peak was located in the first part of the first syllable, whereas in Q2 words, it was located in the second part of the syllable. The sentence-medial peak was 310 Hz, preceded by a rise from 215 Hz and followed by a fall to 210 Hz. The corresponding values for sentence-final peak were 210 Hz, 305 Hz and 205 Hz. Therefore, the sentence-final peak value was 5 Hz lower (310 Hz *vs*. 305 Hz) than the sentence-medial peak value. This small difference was introduced to reflect a slight declination (see t' Hart, 1990: 121f; Maeda, 1976; Pierrehumbert, 1980; Asu, 2004 for Estonian), but was probably not perceivable as such. Both the sentence-medial and the sentence-final peak had a rise of 95 Hz and a fall of 100 Hz. In addition, a non-prominent pitch peak with a rise of 10 Hz on the verb (at the beginning of the sentence) was synthesized. Quadratic interpolations were used for connecting the pitch points that were defined in the syllable-beginnings and at the peak locations. The resulting pitch contours are shown in Figure 2.1.

[6] An utterance without subject and with a sentence-initial verb in the first person plural form in Estonian often induces the pragmatic reading of 'Let's go do it!'

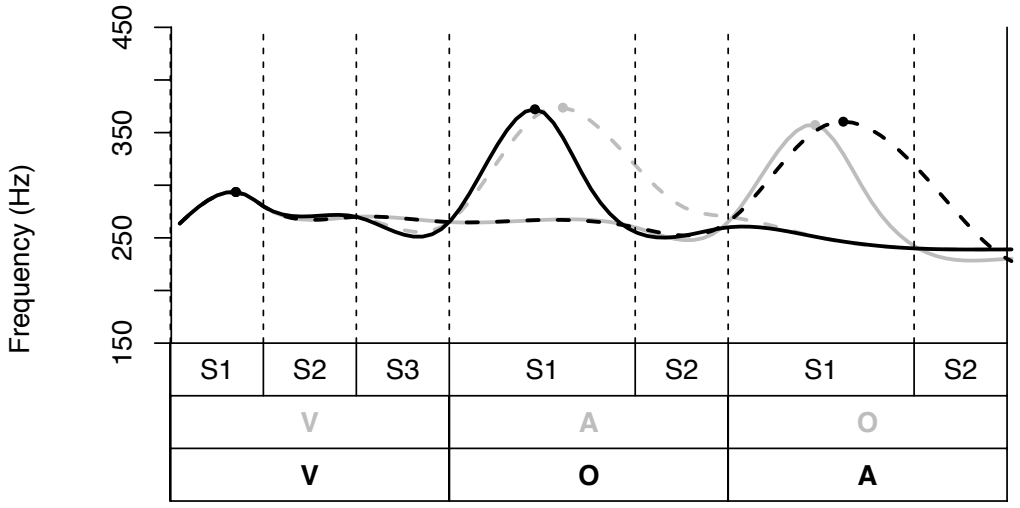

Figure 2.1. Prosodic manipulations of stimulus-sentences. Grey contours represent the word order VAO and black contours the word order VOA. Each word order had a peak located either clause-medially or clause-finally. S1 refers to the first syllable; S2 refers to the second syllable, S3 to the third syllable.

By this manipulation, two additional versions of 24 stimuli that already varied in different word order (VAO *vs.* VOA) were created. With the word order variations and the varied pitch accent placement, the experiment contained altogether 48 different stimuli. There were four versions of each sentence (see the examples in (2.3)):

1) VAO + sentence-final pitch accent (stimulus A)
2) VAO + sentence-medial pitch accent (stimulus B)
3) VOA + sentence-medial pitch accent (stimulus C)
4) VOA + sentence-final pitch accent (stimulus D)

(2.3) Examples of stimuli from (a) to (b) (capital letters show the location of the nuclear pitch accent)

(a) Värvime õuel LAEVA.
(b) Värvime ÕUEL laeva.
(c) Värvime LAEVA õuel.
(d) Värvime laeva ÕUEL.

'Let's paint the boat in the garden.'

To prevent the effect of final lengthening, the tail with a clause consisted of three words *tuli meil mõte* ('we got an idea') with flat contour, which was appended to each stimulus.

2.2.2. Design

The stimuli shown in (2.3) were paired in such a way that each stimulus occurred with all the other stimuli: (a) with (b) (a|b), (a) with (c) (a|c), (a) with (d) (a|d), (b) with (c) (b|c), (b) with (d) (b|d) and (c) with (d) (c|d).

6 combinations of different stimuli were presented with prose texts consisting of 2 to 3 sentences per *context*. There were two types of contexts. The first context (C1) ended with an indirect question about the object (narrow focus on O, see (2.4))[7], the second context (C2) ended with an indirect question about the adverb (narrow focus on A; see (2.5)).

(2.4) The object noun phrase is in focus:

*There was some paint standing on the porch. The weather was nice and we were thinking **what** we could paint in the yard.*

(2.5) The adverb is in focus:

*There was some paint in the porch. Grandfather brought out an old rusty boat, and we were thinking **where** we could paint the boat.*

Each stimulus-pair (a|b, a|c, a|d, b|c, b|d, c|d) was presented with both contexts (C1 and C2). The variation in sentence position (final *vs.* medial) depended on with which context the stimulus sentence was presented. For example, in sentence (2.3a), if presented with the context in (2.4), the word in focus was sentence-final; if presented together with the context in (2.5), the word in focus was sentence-medial.

Nuclear pitch accent either occurred in the word in focus or not: if sentence (2.3a) was presented with a context in (2.4), then the word in focus carried nuclear pitch accent; if (2.3b) was presented with a context in (2.4), the word in focus did not carry nuclear pitch accent.

The combination of 2 word orders, 2 pitch accent placements, 3 occurrences in 6

[7] The Estonian versions of all the stimuli and contexts are provided in Appendix 1.1

stimulus pairs and 2 contexts resulted in 24 conditions for each sentence. The 12 combinations of context and stimulus pair (2*6) were presented with 12 different sentences, which altogether formed 144 experimental items. In order to reduce fatigue, the 144 items were presented in two groups of participants, so that each participant saw only 72 items.

The context was displayed as written text on the screen; stimulus-sentence could be listened by clicking the button on the computer screen. Each stimulus in the pair could be heard maximally three times. The participant could proceed at his own pace, free to make a pause whenever necessary.

The task was to decide whether the first or the second member of the stimulus-pair was compatible with the context displayed on the screen. See Figure 2.2 for the experimental screen.

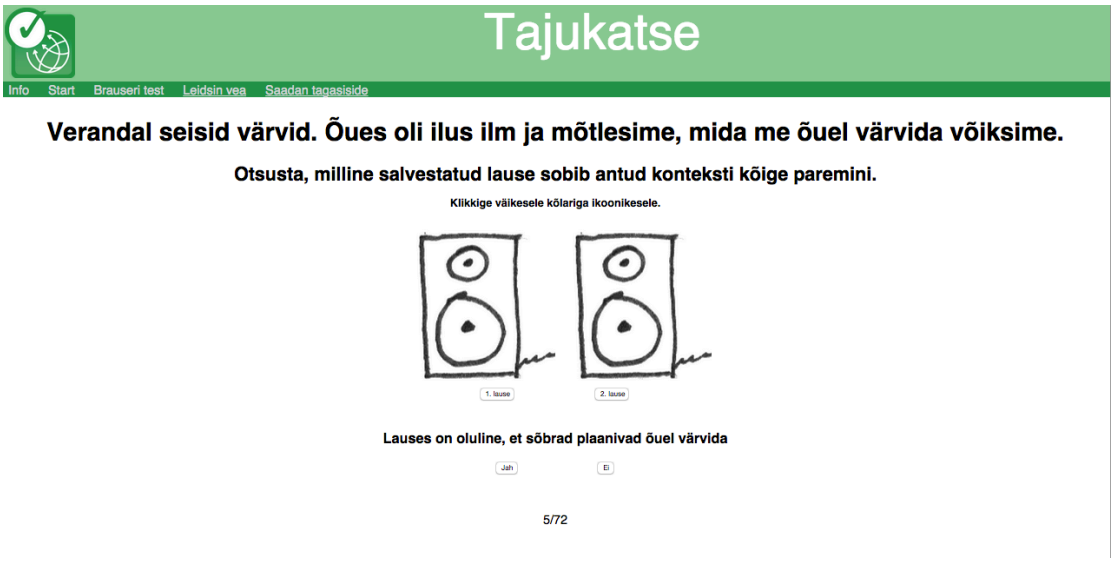

Figure 2.2. An example of the experimental screen. The first row is the context (see the translation in 2.4). The second row is the task: "Decide which of the recorded sentences fits best to the context." The speakers are the buttons for listening the recording, small buttons below them are the buttons to submit the answer. The last row with a text is a comprehension question.

2.2.3. Participants

23 naïve listeners in the age range of 19–56 years participated in the experiment. The listeners originated from different dialectal areas all over Estonia: six from Northern Estonia, four of them from Tallinn; eight from Southern Estonia, four of thme from

Tartu; six from Western Estonia, three of them from Pärnu; two from Central Estonia (Järvamaa) and one from Eastern Estonia. The experiment was carried out with a web interface Percy (Draxler, 2011) and the subjects could take the task from home. The experiment lasted for about 30 minutes and the participants were asked to finish the experiment in one go. The subjects were paid for their participation.

2.3. Results

The experimental design presented above tested the effect of the position and accent on the listener's responses elicited in the forced choice congruence-matching task. The fixed effect position had two levels: sentence-medial (medial) and sentence-final (final); similarly the fixed effect of accent was two-level factor: accented (+acc) and unaccented (-acc). All the statistical analyses were carried out with generalized linear mixed models as a method available in lme4 package (Bates et al, 2012) in R ("R Development Core Team", 2014). The dependent variable was the listener's response encoded as 1 or 2 (either left or right member of the stimulus-pair a|b, a|c, a|d, b|c, b|d, c|d). Random factors were the item (12 different sentences) and the listener (23 participants).

Two types of analysis were conducted. The first-step analysis of all the stimulus-pairs demonstrated how frequently stimulus X was picked. The question is which factor affected the frequency of response the most: accent or position of a word in a sentence. If the stimulus A in the stimulus-pair a|b was chosen, then it was matched with the context written on the screen (response = 'yes'). While matching (a) with the context the stimulus (b) was rejected (response = 'no').

Figure 2.3 shows the number of matches for each stimulus across all the stimulus-pairs.

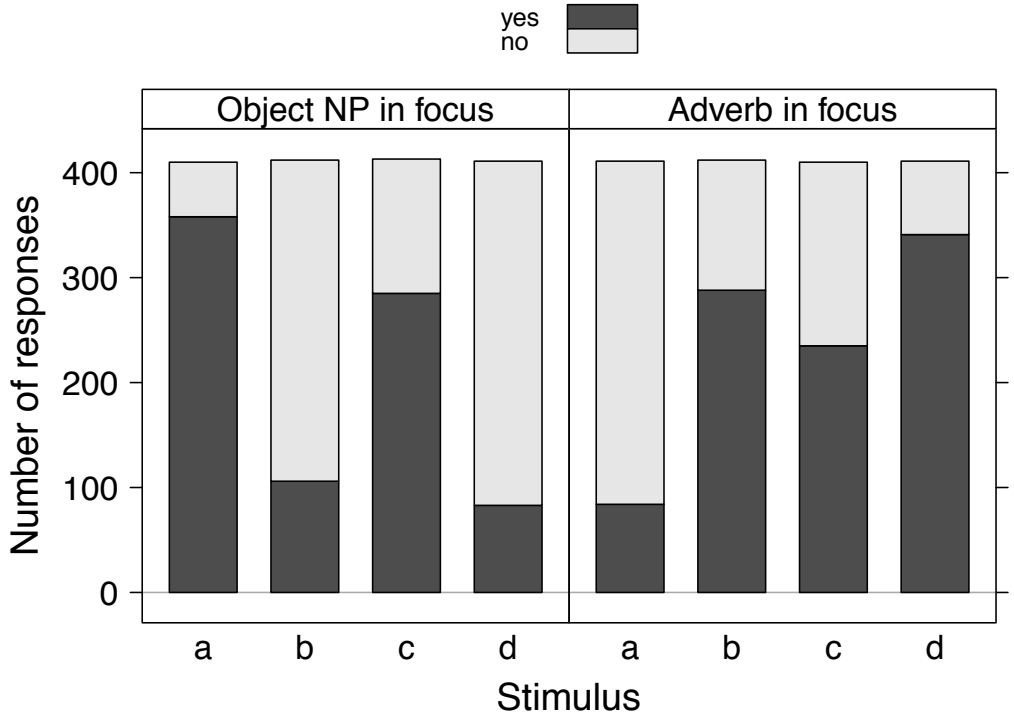

Figure 2.3. The number of responses from the congruence-matching task. a, b, c, d are the stimuli, examples are presented in (2.3), the y-axis shows the frequency whether the stimulus was picked or not across all the stimulus-pairs.

In Figure 2.3 it can be observed that the listeners most frequently preferred stimuli (a) and (c) together with the context of object in focus (C1) and stimuli (b) and (d) together with the context of adverb in focus (C2). Observe that in the context of C1 the stimulus (a) (*Värvime õuel LAEVA*, 'Let's paint the BOAT in the yard') was preferred, whereas in C2 the stimulus (b) (*Värvime ÕUEL laeva*, 'Let's paint the boat in the YARD') was preferred. This cleary demonstrates that an accent shift changes the focus of a sentence in Estonian.

Figure 2.4 plots the number of responses in relation to the position and accent.

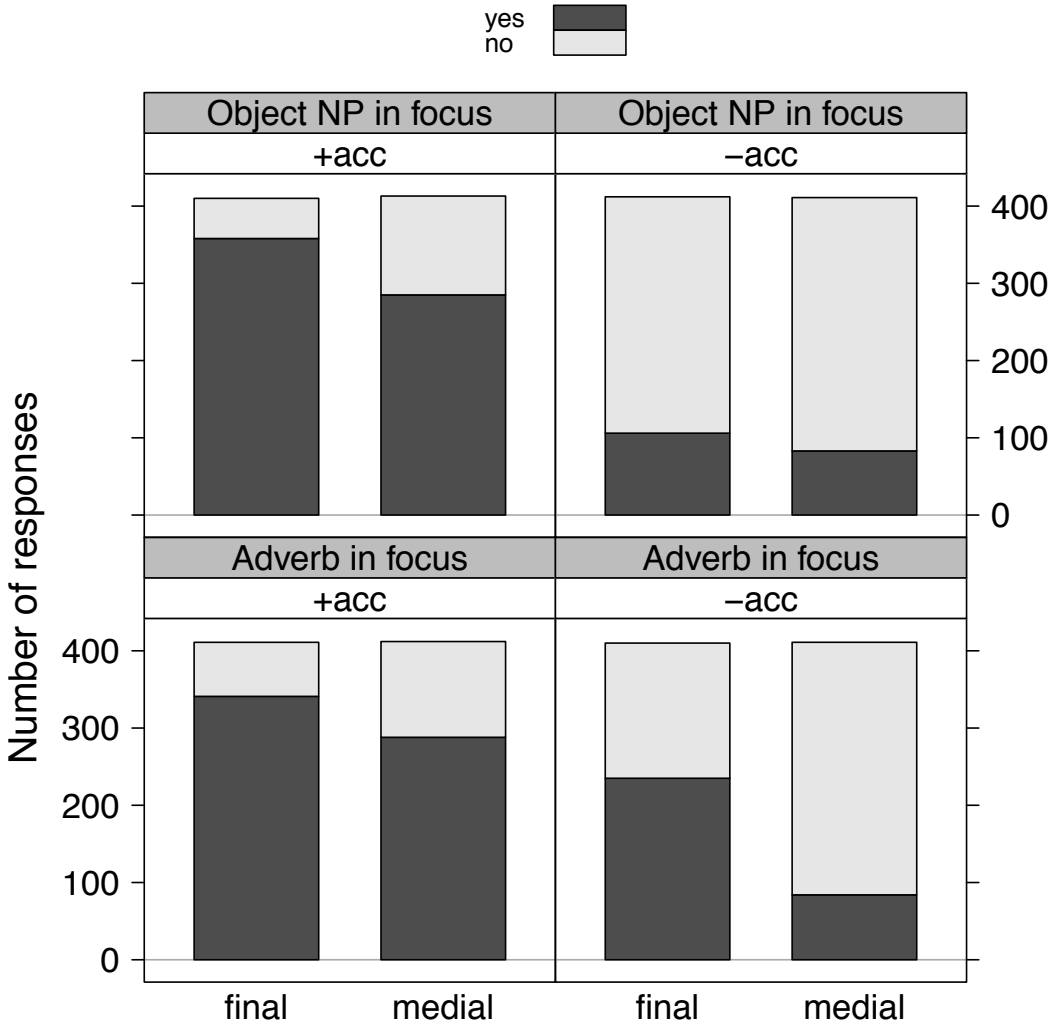

Figure 2.4. The number of responses in relation to position (final means that the word in focus was sentence-final, medial that it was sentence-medial) and accent (+ acc means that the word in focus had pitch accent, -acc means that the word in focus had no pitch accent).

Generalized linear mixed models with the dependent variable response, with fixed effects context (object in focus (C1) *vs.* adverb in focus (C2)), position (final *vs.* medial) and pitch accent (no pitch accent *vs.* pitch accent) and with random variables subject and item showed a significant interaction between the context, position and accent ($\chi^2[1] = 38.11$, $p < 0.001$). Post-hoc Tukey comparisons demonstrated that the final and medial position are significantly different in accented words in both contexts ($p < 0.001$), in the unaccented word in context C2 ($p < 0.001$), but not in C1. Accented

and unaccented are significantly different for the final and for medial position in all contexts ($p < 0.001$). C1 and C2 differed for the sentence-final and unaccented word in focus ($p < 0.001$), but not for the sentence-final accented or for sentence-medial position.

Figure 2.4 shows that position and accent strongly interacted with each other: the frequency of *yes*-matches was the highest when the sentence-final position and the pitch accent coincided with each other. The frequency of *yes*-matches was the lowest when the sentence-final position did not coincide with the pitch accent. If the word in focus was unaccented but varied in position, then there was no difference in response frequency between the sentence-final and sentence-medial position, both are considerably below the chance level. Observe in Figure 2.4 that the unaccented and sentence-final position is exceptional for the adverb in focus (C2). It elicited *yes*-matches slightly above the chance level.

Thus, the results show that the listeners preferred the word in focus to be pitch-accented. However, how did the listeners decid at the point where both of the stimuli in the stimulus pair had a sentence-final and sentence-medial position either accented or unaccented? For this purpose the analysis of response frequencies within a stimulus-pair was conducted.

The second-step analysis poses the question whether the *difference in factor* influenced the distribution of responses in the stimulus-pair. First, see the example in (2.6) where C is the context; (a) and (b) are the stimuli among which the listener had to choose.

(2.6) C: *There was some paint standing on the porch. The weather was nice and we were thinking* ***what*** *we could paint in the yard.*

a. Värvime	õuel	LAEVA.
let's paint	in the yard	the boat
'Let's paint the boat in the yard!'		
b. Värvime	ÕUEL	laeva.
Let's	in the yard	the boat
'Let's paint the boat in the yard!'		

In example (2.6) it can be seen that the context C focuses on the object noun phrase *boat*. In (2.6a) the word *laeva* ('boat') is pitch accented, whereas in (2.6b) it is unaccented. This means that in this stimulus-pair a|b the stimuli differed in accent, but

not in position. In the following we investigate how the difference in factor influenced the responses.

Figure 2.5 presents the number of responses as a function of stimulus type (a, b, c and d) across the 6 stimulus pairs presented in two contexts (object in focus (C1) *vs.* adverb in focus (C2)).

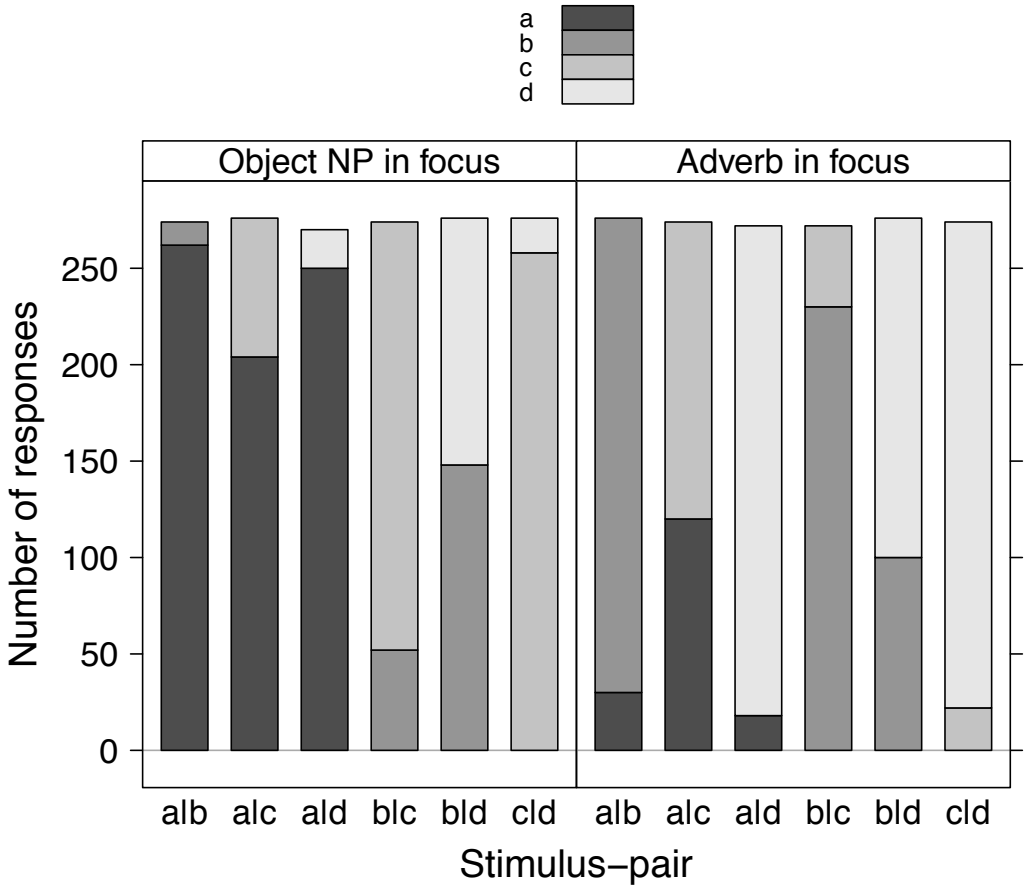

Figure 2.5. The number of responses as a function of the stimulus pair. a, b, c and d are the stimulus sentences (see the examples in (2.3)), x|y shows among which stimuli the participants had to choose while matching the stimulus sentence with the context.

In Figure 2.5, the context C1 is most frequently judged as congruent with stimulus (a) in the stimulus-pairs a|b, a|c and a|d, whereas in pairs b|c, c|d the preferred stimulus is the stimulus (c). In the pair b|d, the distribution of responses between the stimuli is at the chance level for (C1).

In the stimulus-pair a|b in context C1, the stimuli differed in accent but not in position (see demonstration in example 2.6). The first left bar in Figure 2.5 shows that if

there was no difference in position, the listeners matched the stimulus (a) to the context C1. In the stimulus-pair a|c the stimuli differ in factor position but not in factor accent (see (2.3) for reference). The distribution of responses in the second left bar shows that the stimulu (a) was best suited for the context C1. In the stimulus-pair a|d the stimuli in stimulus-pair differed in both factors and this comparison is strongly biased towards the stimulus (a), because it combined both means of focus expression – the pitch accent and the sentence-final position, whereas the stimulus (d) lacked both of them (see examples in (2.3) for reference).

The stimulus-pair b|c differed again in both factors, but not in the same way: in stimulus (c) the sentence-medial object was pitch-accented, whereas in (b) the sentence-final object was unaccented. In Figure 2.5 it can be observed that people preferred stimulus (c) significantly above the chance level. The stimulus-pair b|d differed in position but not in accentuation. Both stimuli are unaccented but the object was sentence-final in stimulus (b). The distribution is at the chance level in Figure 2.5, which shows clearly that the listeners could not decide on the basis of the sentence-final position where the focus of the sentence was. The stimuli in stimulus-pair c|d differed in accentuation but not in position and the distribution of responses in the sixth bar from the left in Figure 2.5 shows that listeners detected the word in focus on the basis of pitch accent.

The results of the analysis of context C2 (where the adverbial was in focus, see the example in (2.5)) are similar to the results of context C1. Interestingly, the responses of stimulus (c) occur slightly above the 50% chance level in stimulus-pair a|c. The pair a|c differed in position but not in accent. This effect was not seen in the context of C1. This suggests that the sentence-final position might be relevant for the adverb but not for the object noun phrase.

Figure 2.6 presents the responses discussed in Figure 2.5 in relation to accent (A) and position (P) and the presence or absence of a difference in these factors, *yes* (Y) and *no* (N) respectively. If the stimuli in the stimulus-pair differed in accent, the pair was encoded ‘accent-yes’ (AY), if the stimuli did not differ, the pair was encoded ‘accent-no’ (AN). If the stimuli presented in the stimulus-pair differed in position, then the pair was encoded ‘position-yes’ (PY) and, if they did not differ, the pair was encoded ‘position-no’ (PN).

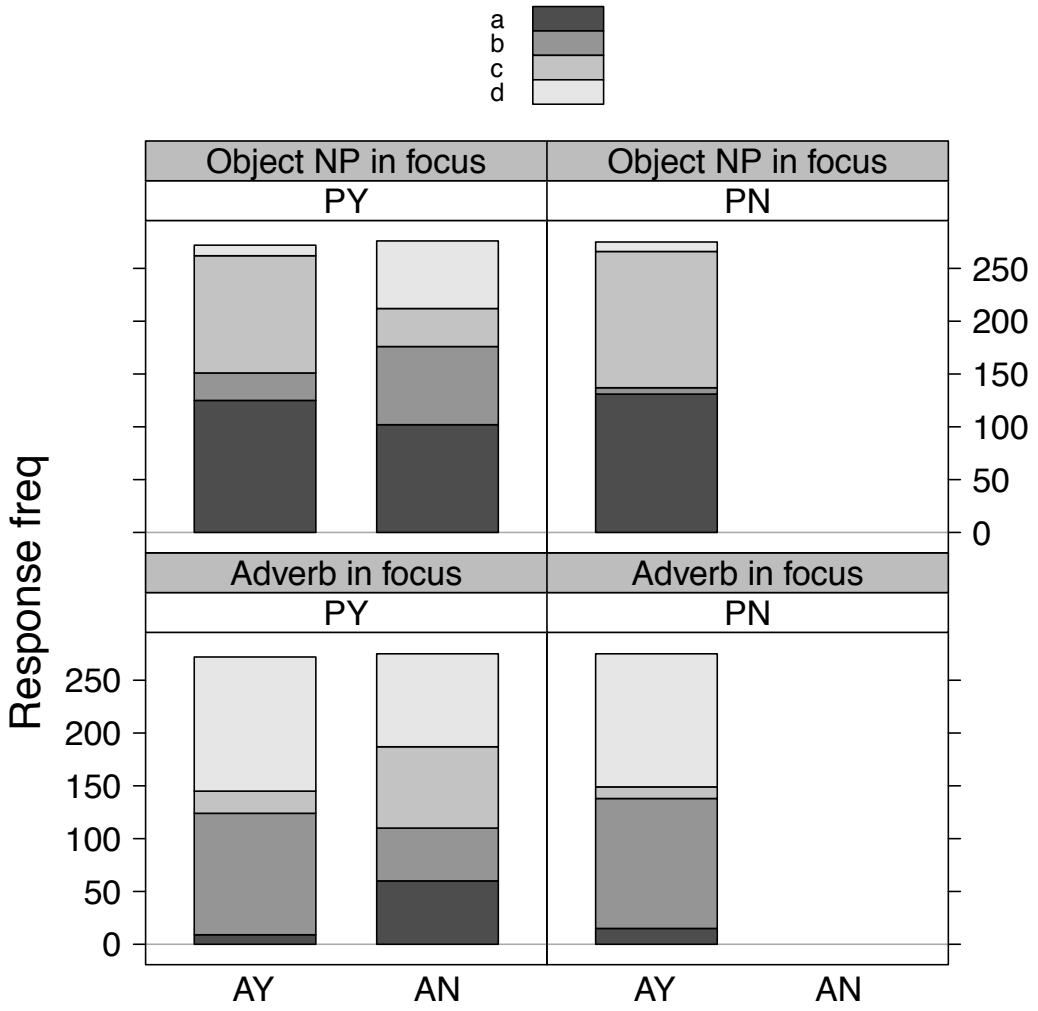

Figure 2.6. The number of responses in relation position (P) and accent (A). 'Yes' (Y) or 'No' (N) encode whether the stimuli presented in the pair differed in the factor or not: PN and AY means that the stimuli presented in the pair had a word in focus in different sentence positions (either medial or final) but the accent placement did not differ.

In Figure 2.6 it can be observed that if the stimuli differed in accent as well as in position (the bar in the middle on the top and bottom panel), then the listeners were very close to chance level in deciding whether there is a focus in the stimulus or not. If the stimuli differed in position but not in accent, then the listeners chose either stimulus (a) or (c) in context C1 (the object in focus) and either (b) or (d) in context C2 (the adverb in focus). The results were the same, if the stimuli in pair differed in accent, but not in position. All the pairs had stimuli differing in either accent or position (therefore, the fourth bar in Figure 2.6 'is missing').

As a next step of the analysis the responses (a, b, c and d) were converted into predicted responses (*p*) and unpredicted responses (*q*) in relation to the hypothesis put forward in section 2.1. The hypothesis stated that the pitch accent cues focus. *p* is the stimulus-pair, in which the listener chose according to the hypothesis, which means that the stimulus where the word in focus was pitch-accented was picked. *q* is the stimulus-pair where the listener chose against the hypothesis, which means that a word that was not pitch-accented or was at the end of the utterance was picked.

In respect to the hypothesis, stimuli (a) and (c) are predicted responses in the context of C1, and (b) and (d) are predicted in the context of C2, because in these stimuli, the word in focus carries a pitch accent (see the examples in (2.3)). Post-hoc Tukey tests showed that the distribution of *p*-s did not differ in two contexts (C1 and C2). For this reason the two data sets were collapsed (see Appendix 2.1 for further details). Thus, Figure 2.7 shows *p/q*-distribution for the stimulus-pairs that either do (Y) or do not (N) differ in accent (A) and position (P).

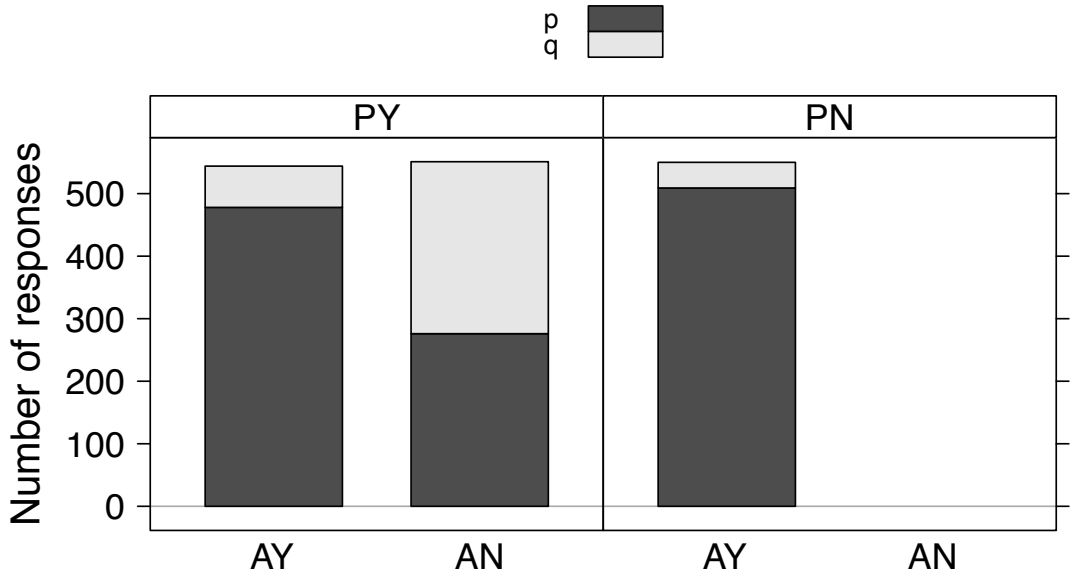

Figure 2.7. The number of responses in relation to position (P) and accent (A). 'Yes' (Y) or 'No' (N) encode whether the stimuli presented in the pair differed in factor or not. p means that the expected response in the condition is the stimulus-sentence that has pitch accent on the word in focus.

Generalized linear mixed models with the dependent variable response frequency (*p/q*-distribution), with fixed effects accent (with difference (Y) or without difference (N)) and position (with difference (Y) or without difference (N)), and with

random effects subject and item showed a significant interaction between accent and position ($\chi^2[2] = 392$, $p < 0.001$). Post-hoc Tukey tests were carried out. AY (difference in accent) and AN (no difference in accent) are significantly different for PY (difference in position) ($p < 0.001$). The difference between PY (difference in position) and PN (no difference in position) for AY (difference in accent) is slightly significant ($p = 0.09$).

The first bar in Figure 2.7 presents the control situation, where the participants picked between stimuli (a) and (d), where the only possibly acceptable answer was stimulus (a). The second and third bars in Figure 2.7 are most interesting for the hypothesis.

When the two stimuli differed in position but not in accent (AN-PY, second bar), then the number of predicted responses was at the chance level. This condition involved, for example, the comparison of stimuli (a) and (c). The word in focus was accented in both stimulus-sentences, but in (a) the word in focus was sentence-final, whereas in (b) it was sentence-medial in C1. The results in Figure 2.7 show that if the word in focus was accented, the sentence-final position was not preferred against the sentence-medial position. Thus, the results clearly show that if the accentuation of the word in focus was unvaried, then the sentence-final position was not enough to cue the focus in a stimulus-sentence.

If the stimuli differed in accent but not in position (AY-PN), then the number of predicted responses was significantly above the chance level. This condition involves the comparison of stimuli (a) and (b). The word in focus is sentence-final in both stimulus-sentences, but in (a) the word in focus is accented, whereas in (b) it is not. In Figure 2.7 it can be seen that the listeners picked the pitch-accented word, if they needed to choose between the sentence-final unaccented and accented word.

2.4. Discussion

The experiment with varying location of a word (medial, final) and varying accent (accented, unaccented) and varying focus (either object or adverb in focus) was carried out in order to investigate the strength of sentence-final position and pitch accent as a focus cue.

The results showed that the word in focus was preferred to be pitch-accented sentence-finally as well as sentence-medially. When the stimuli were paired so that the word in focus was at the end of the sentence and carried the pitch accent in the first stimulus but not in the second stimulus, the listeners preferred the sentence-final position by far. Thus, the sentence-final position is the preferred focus cue, if it is accompanied with the pitch accent. When the stimuli were paired so that the word in focus carried the pitch accent, but was sentence-medial in the first stimulus and sentence-final in the second one, then the listeners could not decide which word is in the focus.

Hence, the experiment showed that intonational prominence is the main cue for focus in Estonian. Sentence-final position cues focus only together with nuclear pitch accent. Moreover, if the position of nuclear pitch accent was changed from the end of the sentence to the middle of the sentence, then it changed the focus of the sentence as well (see Figure 2.3 or 2.7 for reference). Thus, in Estonian, nuclear pitch accent signals focus of a sentence. The sentence-final position did not cue focus.

Our result differs from Finnish, where the position cues focus for some type of constituents (Vainio & Järvikivi, 2006). However, a slight tendency similar to Finnish emerged: the sentence-final position caused perception of focus on the unaccented sentence-final adverb in about 62% (see Figure 2.4 and analysis there). Thus, it appears that in Estonian, if the adverb is sentence-final instead of the object noun phrase, sentence-final position cues focus to some degree.

There are three potential explanations for the strong effect of intonational prominence in results. The first explanation draws on the earlier observations of Estonian. As discussed in the introduction (section 2.1), syntactic focus position (sentence-initial as well as sentence-final) is frequently accompanied by intonational prominence in Estonian. Word order investigations have found that the sentence-initial and sentence-final position both are important for focus in Estonian (Tael, 1988; Erelt et al, 1993; Lindström, 2002; 2004; 2006) while phonetic studies (Eek, 1983; Asu et al., 2009) find that the tonal cues associated with quantity occur in the phrase-final and phrase-initial position, but not in the phrase-medial position. Results of the perception experiment demonstrate a strong interaction between sentence-position and intonational prominence, and strongly suggest that either sentence position together with pitch prominence or just pitch prominence are strong focus cues for Estonian listeners.

The second explanation draws on the design of experimental materials. A relevant factor not considered in the design was the grammatical function of nominal constituents. Vainio and Järvikivi (2006) had stimuli with two adverbs; this experiment varied the location of an object and an adverb. Object noun phrase is the main constituent of the sentence while an adverb is usually an optional constituent. This grammatical difference between the sentence constituents may have suppressed the effect of a sentence-final position in focus perception. Several other studies (Gussenhoven, 1983a; Truckenbrodt and Darcy, 2010) have noted, for example, that the difference between main and optional constituents (sentence arguments *vs.* modifiers) affects sentence intonation. The type of grammatical function (or semantic role) might affect the perception of sentence prominence in a way that object noun phrase is congruent with different kinds of foci (see e.g. examples 1.8 and 1.9 in Introduction).

Another explanation is related to the theory of word order in Estonian. Erelt et al. (1993) observe that the sentence-final information is *rheme*, which means that this is the part of the proposition that states something new about the referent. Therefore, sentence-final position is interpreted as a focus cue. However, Tael (1988: 10–11) observes in her terms that for some constituents, the sentence-initial position is more 'emphatic'. In Lindström (2006: 879), it is explicitly stated that the sentence-final position is the default ('unmarked' in her terms) focus, whereas the sentence-initial position is more specific ('marked' in her terms). Therefore, theoretically the sentence-final position might somehow be a weaker cue for focus position than the sentence-initial position. This may explain why the sentence-final position did not have any influence on focus perception.

The third explanation is concerned with the phonetic characteristics of the stimuli. The only prosodic cue that was manipulated in the stimuli was the F0 contour; the duration, intensity and vowel quality of the original recordings were preserved. The original recordings were in two versions: either with an accent on phrase-medial or on phrase-final constituent. The F0 contour was either enhanced (enlargement of the F0 expansion) or just stylized. Therefore, the participants might have been forced to make use of the pitch cue in the stimuli, in which the intensity and duration was unvaried. To us, the next logical step of the investigation would be to see whether the speakers would expand the F0 curve on the word in narrow focus in a similar way like it was expanded in the stimulus-utterances of the current study.

The main outcome of the study is in accordance with many other studies and theoretical claims. First, Keller and Alexpoulou (2001) showed that pitch accent could convey focus of a sentence independently from word order or morphological markers and even override the information-structural implication that the sentence position or the morphological marker had. Second, Skopeteas et al. (2009) shows by the theoretic analysis of their empirical data that if there is any ranking relation posed, then the prosodic means are more important for focus marking than the syntactic means.

2.5. Conclusion

A perception study was run to test whether focus perception is connected to sentence-position or to pitch accent in Estonian. The task for naïve listeners was to decide within a particular context which of the two utterances had congruent focus with context. The results showed that the sentence-final position alone was a weak cue for focus and that the position interacted strongly with intonational prominence. The intonational prominence was to be the strongest cue. The result of the study differs from the results of Finnish, which is a closely related language, and demonstrates how important detailed investigations of individual languages are. The study provides evidence that it is not possible to extrapolate from a model of a language to a model of another, even if the languages are closely related (e.g. from Finnish to Estonian or from English to German).

The study provides additional data that pitch accent is relevant for focus also in languages that cue focus with a position in a sentence. In the next chapters, the phonetics of pitch accent on a word in focus is going to be investigated. There is another position in the sentence that is information-structurally relevant in Estonian – the beginning of a sentence. It might be that the sentence-initial position takes stronger effect on production and perception of focus prosody than the sentence-final position.

3. Broad and narrow focus in Estonian

Abstract

The study investigates and re-examines the phonetics of narrow focus in Estonian speech. A perception study in chapter 2 showed that pitch prominence is perceived as focus in Estonian. Conversely, Sahkai et al. (2013b) report that focus in Estonian is produced with prosodic prominence but not with F0 expansion. If listeners perceive pitch accent as focus, does this imply that they also employ pitch prominence for focus production? In addition, a question whether a word order consisting of structural focus position (O_FVS) can be produced with broad focus intonation. Vainio and Järvikivi (2007) have shown that L1-Finnish speakers produce intonation with declinating pitch peaks on all the nominal constituents (broad focus intonation) in sentences with sentence-final focus position, if stipulated by the previous linguistic context. Thus, the study investigates in a manner similar to Vaino and Järvikivi (2007) whether OVS word order in Estonian can be produced with 'neutral' broad focus prosody.

A speech production study with a speech elicitation task was carried out. L1-Estonian speakers had to reply to questions that were designed to elicit answers with broad and narrow focus on either subject or object noun phrase. The answers were uttered following the appearance of a series of pictures shown on the screen. For the acoustic effects of narrow focus, the size of the F0 excursion on the basis of several parameters was evaluated. In addition, peak difference between the peaks found on the noun phrases was calculated.

Pitch prominence interacted with grammatical function of the referring expression. Narrow focus was produced with greater pitch prominence on the subject but not on the object noun phrase. Peak difference demonstrated that OVS word order is preferred with nuclear pitch accent in the beginning of the phrase and not with 'neutral' broad focus intonation. The experiment showed that Estonians use pitch prominence to cue focus that is stipulated by the previous linguistic context. However, word order (e.g. OVS) might determine the location of pitch prominence for some types of linguistic contexts (e.g. broad focus).

3.1. Introduction

As described in chapter 2, pitch accent in a sentence caused a strong perception of focus in Estonian. This result serves as a basis for investigating the phonetics of sentence accent in a speech production experiment.

Pitch prominence has been attested to be the main acoustic cue for narrow focus in intonation languages such as English and German (Cooper et al., 1985; Eady & Cooper, 1986; Swerts et al., 2002; Baumann et al., 2006; Féry & Kügler, 2008; Breen et al., 2010). A phonetic realisation of a pitch peak has been shown to vary in height as well as in the extent of the fall (Baumann et al., 2006, Breen et al., 2010; Swerts et al, 2002; Féry & Kügler, 2008) and in alignment (Pierrehumbert & Steele, 1986; Silverman and Pierrehumbert, 1990; Kohler, 1987ab; 1991; Prieto et al., 1995; Ladd and Morton, 1997). Moreover, a higher peak, steeper F0 range and a later peak (Ladd & Morton, 1997; Kohler, 1991) are reported to cause perception of a stronger accent. Thus, the gradual variation in pitch could be interpreted categorically and the systematic variation of F0 may result in categorical difference between *weak* and *strong pitch prominence* (Terken, 1991; Ladd & Morton, 1997) that might differentiate narrow focus from broad focus.

However, studies have shown that focus influences in addition to F0 also intensity, duration and vowel quality (Fry, 1955; 1985; Beckman & Edwards, 1994; Breen et al., 2010). Empirical data shows that words in focus are significantly longer and with higher intensity maximum than the corresponding words in broad focus. The studies on corpus data conclude that duration and intensity are much more stable cues in focus production than F0 (Turk & Sawusch, 1996; Kochanski et al., 2005; Cole et al., 2011).

Sahkai et al. (2013ab) have found that L1-Estonian speakers use prosodic prominence for focus (but not predominantly F0 expansion). In Sahkai et al. (2013b), duration was strongly affected by narrow focus, whereas F0 range was not. The perception study in previous chapter showed, however, that L1-Estonian listeners perceived pitch accent as focus. If the listeners perceive pitch accent as focus, does it also imply that they use pitch prominence consistently also for focus production: do they produce stronger pitch prominence in narrow focus than in broad focus? Based on the results presented in chapter 2, it is predicted that narrow focus is produced with a

stronger pitch prominence than broad focus. The prediction is supported by numerous studies of intonation languages (Cooper et al., 1985; Eady & Cooper, 1986; Swerts et al., 2002; Baumann et al., 2006; Féry & Kügler, 2008; Breen et al., 2010).

The prosodic means are not the only means to highlight focus in Estonian. A few studies suggest that there are some sentence positions that lend themselves to focus interpretation (Tael, 1988; Erelt et al., 1993; Erelt, 2009; Lindström, 2004; 2006). The observations suggest that the 'normal' default focus is located at the end of the sentence, whereas the 'special' more emphatic focus is in the beginning of the sentence (Erelt et al., 1993; Lindström, 2006). An interpretation relevant to the study is that the word in narrow focus can either be located sentence-initially or sentence-finally, but the focus in the beginning of the sentence is impressionistically stronger than the focus at the end of the sentence. In the experimental study by Sahkai et al. (2013a), the speakers did not use the word order permutations for identifying focus. However, the different word orders are quite frequent in the corpora of written language and of spontaneous speech (Tael, 1988; Lindström, 2002; 2004). This fact serves as a basis for investigating the effects of word order in Estonian sentence intonation.

Finnish, a typologically related language (Abondolo, 1998) is claimed to use word order inversions for the expression of focus similarly to Estonian, but it is also known to use prosodic means for focus marking (Välimaa-Blum, 1988; 1993; Vainio & Järvikivi, 2006; 2007). Martti Vainio and Juhani Järvikivi (2006; 2007) have studied the interaction between word order and prosody for communicating focus in Finnish. In Vainio and Järvikivi (2007), they carried out a speech production study where they investigated the production of pitch prominence in two types of sentences embedded into a broad focus context. The sentences were elicited as responses to the question *What happened?* The difference between the two types of sentences was that in the first type (e.g. *Menemme Jimille* ***laivalla***, 'It is by boat, we go to Jimi') the constituent in the sentence-final position was in focus due to word order (location adverb in sentence-final position) – the so-called *structural focus*, whereas in the second type (e.g. *Menemme laivalla Jimille*, 'We go to Jimi by boat') there was no constituent in focus due to word order. On the basis of their previous perception study (Vainio & Järvikivi, 2006), Vainio and Järvikivi (2007) predicted that the pitch peak on a word that was in focus due to the sentence-final position (*laivalla*) is lower than in the word *Jimille* that is not in focus position (it is in its 'canonical' position). They expected their speakers to put

more effort into signalling broad focus in the word that is in a structural position of narrow focus. A assumption of this prediction was that the speakers of Finnish differentiate between broad and narrow focus by intonation (as Suomi et al., 2003 have also shown). Vainio and Järvikivi (2007) call their prediction as *prosodic compensation*: the speakers neutralize abstract prominence elicited by the sentence-final position by using prosodic prominence. In other words, the prosodic prominence was expected to compensate for the so-called *structural prominence* in sentences with sentence-final focus position, if they are uttered in the broad focus context. Results in Vainio and Järvikivi (2007) give evidence that the speakers indeed compensate for structural prominence by means of prosody. To put it more simply, the speakers were able to produce a sentence with the sentence-final focus position with broad focus intonation.

The same question as in Vainio and Järvikivi (2007) is going to be investigated in a different sentence type. The aim is to explore possible prosodic compensation in OVS word order, where the object noun phrase is in the sentence-initial focus position. In the speech production experiment, the effect of sentence-initial position in the production of broad focus intonation in Estonian is tested and the question is whether speakers are able to utter OVS word order with broad focus intonation? Theoretically, the focus in the beginning of the sentence is stronger than at the end of the sentence. Therefore, the prediction is that the OVS word order is not possible with broad focus intonation.

3.2. Experiment

The predictions in section 3.1 were tested in a speech production experiment, in which participant was asked to respond to the question heard over headphones following the appearance of a series of pictures.

3.2.1. Materials

Four sentences were constructed as *target sentences*. Target sentences consisted of three two-syllable words with mostly sonorous sounds, see the list in (3.1).

(3.1)

1. *Leena maalis vaala*. (Lena drew a whale.)
2. *Liina liimis raami*. (Liina repaired a frame.)
3. *Meeli hüüdis Loonat*. (Meeli called for Loona.)
4. *Miili kuulis Eevat*. (Miili heard Eva).

The target sentences in (3.1) are all SVO sentences that were reverted to OVS sentences, so that all subject noun phrases and object noun phrases occurred sentence-initially as well as sentence-finally. All the subjects were disyllabic words of quantity two (Q2). The objects were disyllabic words of quantity three (Q3) in the first two sentences (*vaala*, 'whale'; *raami* 'frame') and of quantity two (Q2) in the last two sentences.

The sentences were triggered as responses to questions or assertions called *context* (presented in (3.2))[8].

(3.2)

1. What's up?
2. Somebody drew {repaired|called|heard} a whale {a frame|Loona|Eva}?
3. Leena {Liina|Meeli|Miili} drew {repaired|called|heard} something?

The 2nd and the 3rd context in (3.2) were *wh*-questions that were signalled by intonation. The context was meant to be interpreted so that the listener is asked to whom *somebody* and to whom or what *something* refers. This way of presenting the *wh*-question was chosen due to an intuition – which was confirmed in the pilot experiments – that it is very unnatural to answer the *wh*-question with a full sentence (similar kind of difficulty is reported also in Swerts et al., 2002). The speakers of the pilot study reported that they would prefer to give a one-word-response. If an intonational question like the 2nd and 3rd context was offered as an alternative, then a three-word-response was found to be more natural. The participants suggested that it is possible to imagine an additional context where the speaker is asking for a repetition, because (s)he did not hear the entire sentence in the first place.

[8] See appendix 1.2 for the Estonian versions.

Two word orders (SVO, OVS) crossed with three contexts (broad, object in focus, subject in focus) resulted in 6 conditions for each of the four sentences presented in (3.1). The list of target sentences therefore consisted of 24 items.

In addition to these 24 items, four other three-word-sentences called *filler sentences* were constructed with an object noun phrase (O), a verb (V) and a sentence modifier (adverb (A)). Word order in filler sentences was varied similarly to target sentences (AVO *vs.* OVA). The filler sentences (24 items) were presented in similar contexts as shown in (3.2). In total, the participant was presented with a randomized list of 48 items (24 targets and 24 fillers).

3.2.2. Procedure

The participants together with the experimenter were seated in front of a computer screen in a soundproof booth. The experiment was run as a slide presentation (demo window) in Praat (Boersma and Weenink, 2013). Participants proceeded from slide to slide at their own pace with a mouse click. They were free to make a break or ask the experimenter for a clarification any time they needed.

The context was presented acoustically with the appearance of a new slide. It was played over a computer with internal speakers and additionally written in the upper part of the screen. Listeners were able to listen to the context twice before they produced the target sentence.

The target sentence was displayed as a series of pictures. Each subject, object and adverbial was schematically depicted (see Figure B.2 in Appendix B for sample pictures). The participants were asked to memorize the pictures together with the word forms. The pictures were displayed on the screen in the order of the expected word order: SVO, OVS, AVO, OVA (see Figure 3.1).

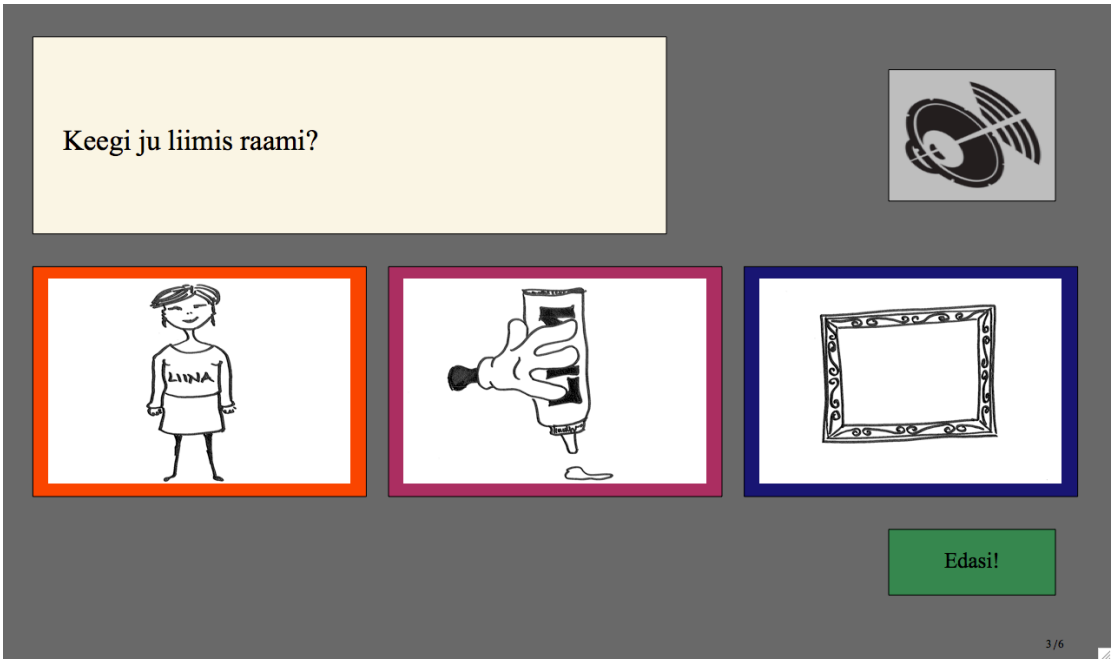

Figure 3.1. Visual display of the experimental item. The context *Keegi ju liimis raami?* ('Somebody repaired the frame?') was written in the upper third of the screen. The button in the upper right corner replays the context as a sound. The target sentence is displayed as a series of pictures: the subject noun phrase *Liina* was framed with orange, the verb *liimis* ('glued, repaired') with dark pink, and the object noun phrase *raami* with dark blue. The participant was supposed to utter a target sentence in an SVO word order with narrow focus on the subject.

Participants were asked to compose a sentence from the sequence of pictures and respond to the context they heard and read. The experimenter asked them to correct themselves, if the word order in an utterance did not follow the order of the pictures.

The procedure took about 20 minutes and consisted of three sessions: the picture memorizing task, training, and the experiment. Utterances were recorded with a head-mounted microphone at a rate of 44 kHz.

3.2.3. Participants

The participants were ten female and seven male speakers (17 altogether) between 22 and 40 years of age (m = 28.2 years). Ten speakers originated from the Northern Estonian dialectal area, three of them from Tallinn; seven speakers came from Southern Estonia, three of them from Tartu. They were either students at the University of Tartu or professionals from Tartu and Tallinn. They did not report any visual or hearing deficiencies. The participants contributed voluntarily.

3.2.4. Analysis

The acoustic analysis of the recordings was carried out with Praat (Boersma and Weenink, 2013) only for high falling pitch excursions (pitch accent type H* or H*L). Due to a small number, 42 instances of low accent (L*) were excluded from the evaluation. Therefore, together with the utterances with incorrect word order and hesitations, 7% of the utterances were omitted.

The F0 contour was manually segmented, relying on perception and visual observation of the F0 track. In the nominal constituents of the sentence (grammatical subject and object), F0 maximum or if a clear F0 maximum was missing, the so-called *elbow* where F0 started an abrupt change either from high to low or from low to high – *plateau offset* (see similar definition also D'Imperio, 2000; Knight, 2003) was determined. See Figure 3.2 for illustration. The plateau offset on the F0 track (grey line) was determined to be where the angle between the red reference line (provided in the Praat editor) and F0 track increased abruptly (vertical dotted line at the time of 0.18). The *onset* of the next plateau was determined where the angle between the reference line (blue in Figure 3.2) and F0 track decreased abruptly (vertical dotted line at the time of 0.29 s).

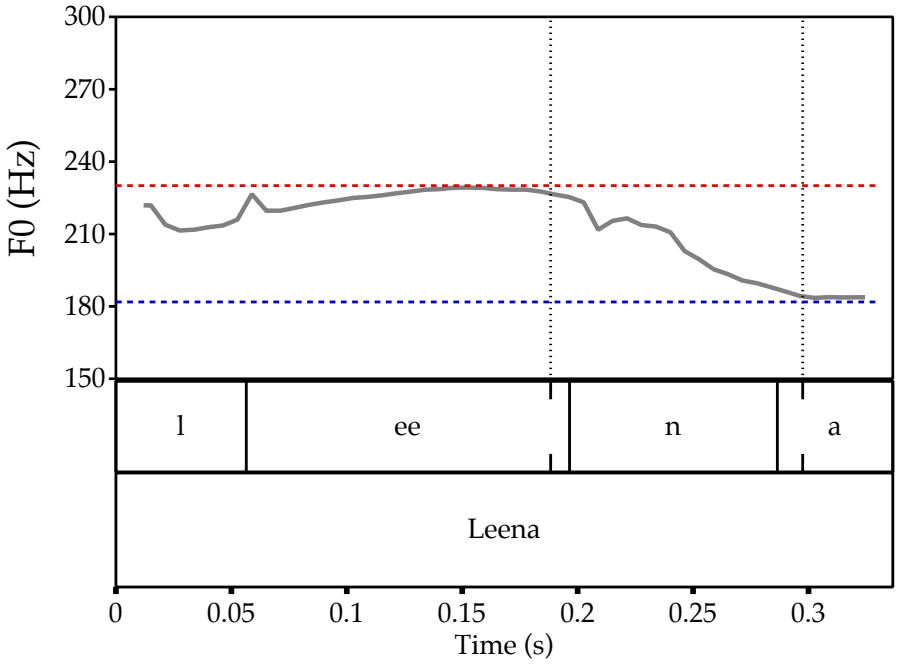

Figure 3.2. Approximate determination of plateau offsets and onsets in the Praat editor window with reference to the horizontal red line provided in Praat (red and blue lines in the Figure).

In this study, the plateau offset is called as F0 turning point (TP) and the onset of the next plateau as F0 target (TF0). The visual determination against the automatic is quite approximate and the variation present in the data is acknowledged. The results of the experiment are discussed on the basis of several tonal variables extracted from the annotation – *F0 peak*, *slope, peak alignment* and *peak difference*.

The *F0 peak* is the F0 maximum as well as the F0 turning point, annotated as TP (Figure 3.2). The peak height was estimated with reference to the mean F0 of the speaker (spMean). The peak values were converted into semitones using the formula in (3.3):

(3.3) Semitone (st) = $12\log(F0/F0_{spMean})$

The speaker mean was calculated as the mean value of all the F0 samples generated by Praat pitch analysis (default time step, pitch floor 75 Hz) from all the uttered items. The conversion of semitone-interval in reference to mean F0 of the speaker was chosen in order to reduce the speaker variability. Based on the earlier results for other intonation languages (Baumann et al., 2006, Breen et al., 2010; Swerts et al., 2002; Féry & Kügler, 2008; Vainio & Järvikivi, 2007), the peak is expected to be higher in narrow focus than in broad focus, if different focus types are signalled with different degrees of intonational prominence.

F0 slope (procedure adapted from t' Hart et al., 1990; Niebuhr, 2007) was defined as a main correlate of intonational prominence: the range of the F0 change was divided by the duration of the excursion. See the formula in (3.4) where the $F0_{TP}$ is the F0 (in semitones) of the F0 maximum or the plateau offset, the $F0_{TF0}$ is the F0 of the F0 onset of the next plateau, here called F0 target (see Figure 3.2 for reference); T_{TP} (in ms) is the absolute time of the turning point (TP), T_{TP} the absolute time of the F0 target.

(3.4) Slope (st/s) = $(F0_{TP} - F0_{TF0}) / (T_{TP} - T_{TF0})$

Slope shows the speef of the F0 change: the faster the change, the steeper the F0 drop (or rise). The advantage of reporting the slope instead of range is that it normalizes for the F0 fall and F0 range of large magnitude that might happen to occur within a long timespan. The slope will be small in this case and reflect the fact that the F0 change was

slow and resulted in an F0 excursion that was rather flat. Figure 3.3 illustrates the relation between the excursion and the slope.

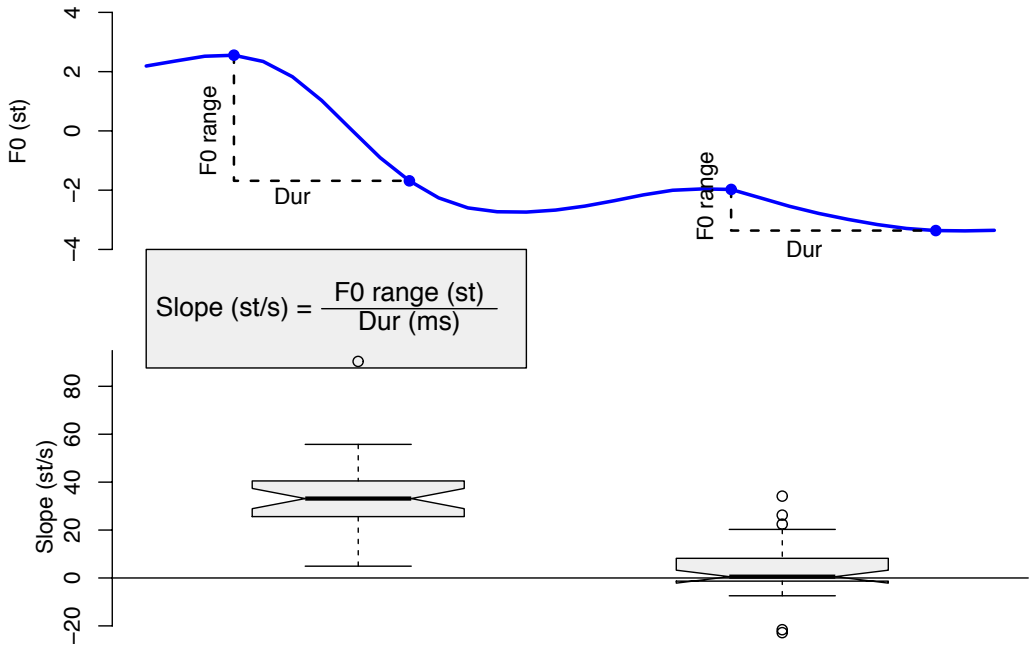

Figure 3.3. F0 excursions and their slopes. The upper part of the figure shows normalized F0 aggregated over an utterance type of *Leena maalis vaala* ('Lena drew a whale') uttered by 17 participants. The bottom part of the figure shows boxplots of the slopes calculated (formula in the grey box) for the two pitch excursions seen in the upper part.

Thus, the slope was included into parameters of intonational prominence together with the peak height. The faster the F0 change, the steeper is the F0 excursion and the greater is the value for slope. Based on the earlier results on other intonation languages (Cooper et al., 1985; Eady & Cooper, 1986; Lieberman, 1960; Couper-Kuhlen, 1984), F0 slope is predicted to be smaller in broad than in narrow focus, if the narrow focus is signalled with stronger degree of intonational prominence.

Peak alignment was defined as the time from the vowel onset to the time of the F0 peak in proportion to the duration of the first stressed vowel. See the formula in (3.5) (procedure adapted from Ratchke & Harrington, 2007; Plüschke, 2013),

(3.5) $T_{prop} = 100(T_{F0}-T_{vOn})/(T_{vOff}-T_{vOn})$

where T_{F0} is the time of the F0 point, T_{vOn} the time of the vowel onset and T_{vOff} the time of the vowel offset. Relying on the earlier claims about peak alignment in relation to narrow focus intonation (Silverman & Pierrehumbert, 1990; Kohler, 1991; Prieto et al., 1995; Ladd & Morton, 1997), the peak was predicted to be later in narrow focus than in broad focus, if narrow focus is signalled with stronger degree of intonational prominence. The peak alignment depends on quantity in Estonian. The investigation of peak alignment in connection to focus type was also expected to reveal an interesting interaction between sentence intonation and word prosody.

For *peak difference*, the difference between F0 peaks was calculated by subtracting F0 (in semitones) of a second peak from the first. The first nominal constituent of the sentence carried the first peak. The second nominal constituent (either grammatical subject or object) carried the second peak. See Figure 3.4 for exemplification.

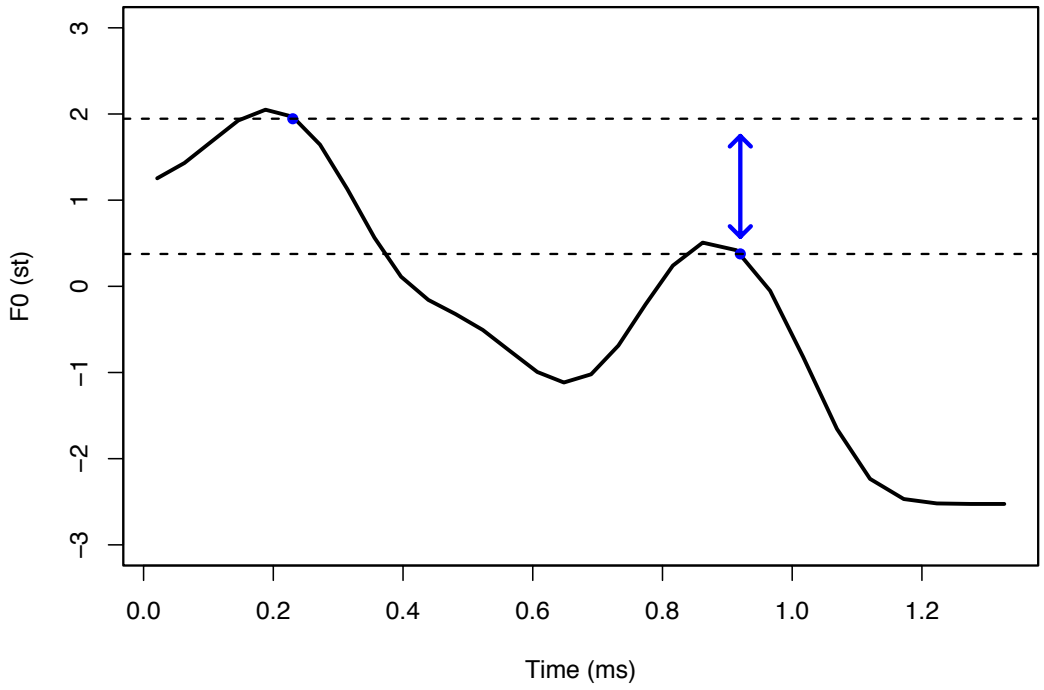

Figure 3.4. Peak difference. Peak difference was calculated by subtracting F0 (in st) of the second peak (or plateau offset (TP)) from the F0 (in st) of the first peak (TP).

In American English, if there is F0 expansion on the word in focus, then F0 of a word in non-focus is compressed (Cooper et al., 1985; Xu & Xu, 2005). In this study, this effect is quantified as peak difference. Peak difference scans the relation between

the peak values – the so-called *topline*. With regard to this, the effect of declination needs to be addressed.

For physiological reasons, F0 drops gradually towards the sentence end and, among other effects, this causes every next pitch peak or valley to be lower than the previous pitch peak (t'Hart and Cohen, 1973; Maeda, 1976, Cohen et al., 1982). The listeners are known to compensate for the downward drifting topline by perceiving the lower peak towards the sentence end as being at the same level of prominence as the previous one (Pierrehumbert, 1979; Terken, 1991). Therefore, there is no reason to expect a significantly positive slope if the word in focus is located at the end of the sentence. In the utterances with sentence-final focus, the peaks are expected to be maximally at the same level or the second peak to be slightly lower from the first one. A positive slope would unexpectedly show a strongly upstepped pitch peak on the final word. Taking this into an account, peak difference for sentence-initial focus is expected to be significantly negative, because the first peak is considerably higher than the second peak. The peak difference for broad focus or sentence-final focus is predicted to be either a zero due to the higher pitch on the sentence-final word in focus, or slightly negative due to the effect of declination. For the word order effect, the peak difference of OVS should be different from the peak difference of SVO in the broad focus context.

For this study, peak height, slope, peak alignment and peak difference are defined as parameters of intonational prominence. The fixed effects of focus type, grammar and sentence position were evaluated on the basis of the 4 dependent variables: peak height, slope, alignment and peak difference. All the statistical analyses were carried out with software R for statistical computing ("R Development Core Team", 2014) as a method available in lme4 package (Bates, et al., 2012).

3.3. Results

Recall from example (3.1) that there were Q2 as well as Q3 words included as object noun phrases in the experimental materials. Table 3.1 lists the analysed target words as a function of grammatical function and quantity.

Table 3.1. List of target words. Quantity is notated in the brackets: Q2 is the second quantity, Q3 the third quantity.	
Subjects (quantity)	**Objects (quantity)**
Leena (Q2)	Eeva**t** (Q2)
Liina (Q2)	Loona**t** (Q2)
Meeli (Q2)	**vaala (Q3)**
Miili (Q2)	**raami (Q3)**

Subject noun phrases were all disyllabic words of Q2, whereas the quantity of objects vary between Q2 and Q3. The first two objects are proper names *Eeva* and *Loona* are Q2 words with a partitive marker *-t* (in bold in Table 3.1). The other two objects are Q3 words with an overlong vowel (in bold in Table 3.1). The effect of quantity prior to any further evaluation was examined. The number of Q3 words was considerably smaller in the data than the number of Q2 words. There were 107 productions of Q3 words and 340 productions of Q2 words.

Figure 3.5 plots the peak height in relation to quantity.

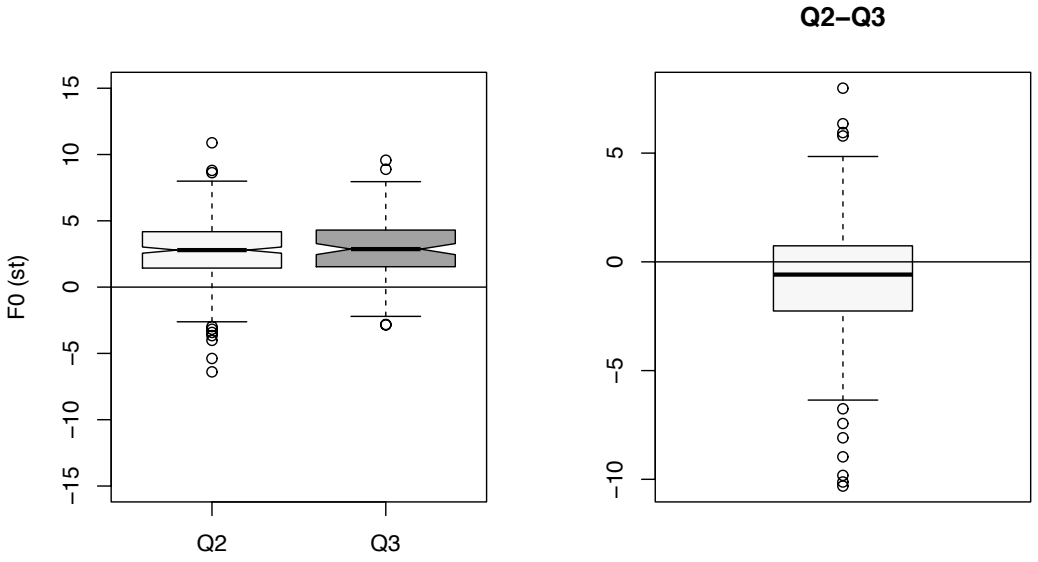

Figure 3.5. Peak height in relation to quantity. On the left: peak height (in semitones) in relation to quantity; on the right: the distribution of peak differences between quantity two (Q2) and quantity three (Q3).

The Repeated Measures ANOVA (RM-ANOVA; as provided in the R package "ez" ("R Development Core Team", 2014)) with dependent variable peak height, random

variable subject and the within variable quantity[9] showed no significant effect of quantity on the peak height. As can be seen in Figure 3.5, the peaks in Q2 and Q3 are at the same height and the differences between the peaks are very close to 0. The data shows that the quantity does not significantly affect the peak height.

Figure 3.6 plots the slope in relation to the quantity.

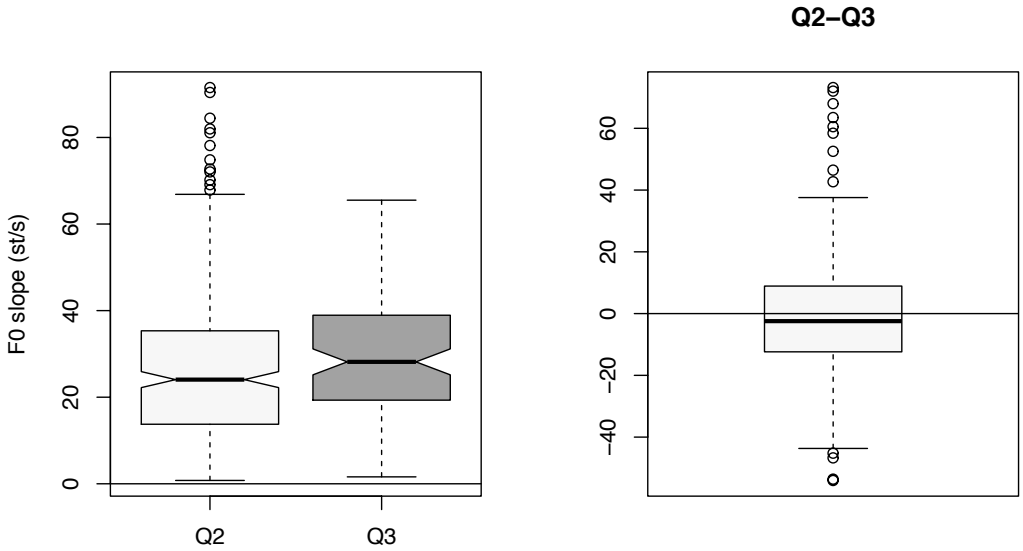

Figure 3.6. F0 slope in relation to quantity. On the left: slope (semitone in second) in relation to quantity; on the right: the distribution of peak differences between quantity two (Q2) and quantity three (Q3).

The RM-ANOVA with the dependent variable slope, random variable subject and the within variable quantity[10] showed a significant effect of quantity on the slope ($F[1,16] = 4.4$, $p = 0.05$). Therefore, the quantity affects the slope. Figure 3.6 shows that the slope is slightly greater in Q3 words than in Q2 words, but the effect is rather small; the distribution of differences is very close to zero at the right-hand plot of Figure 3.6.

Next, the peak alignment will be examined. On the basis of previous studies of Estonian quantity, an early peak for Q3 and a late peak for Q2 were expected (Lehiste, 1960, 1997; Lippus et al, 2013). Figure 3.7 plots the peak alignment in relation to the quantity.

[9] ezANOVA(peakData, .(peak), .(subj), .(quantity))
[10] ezANOVA(slopeData, .(slope), .(subj), .(quantity))

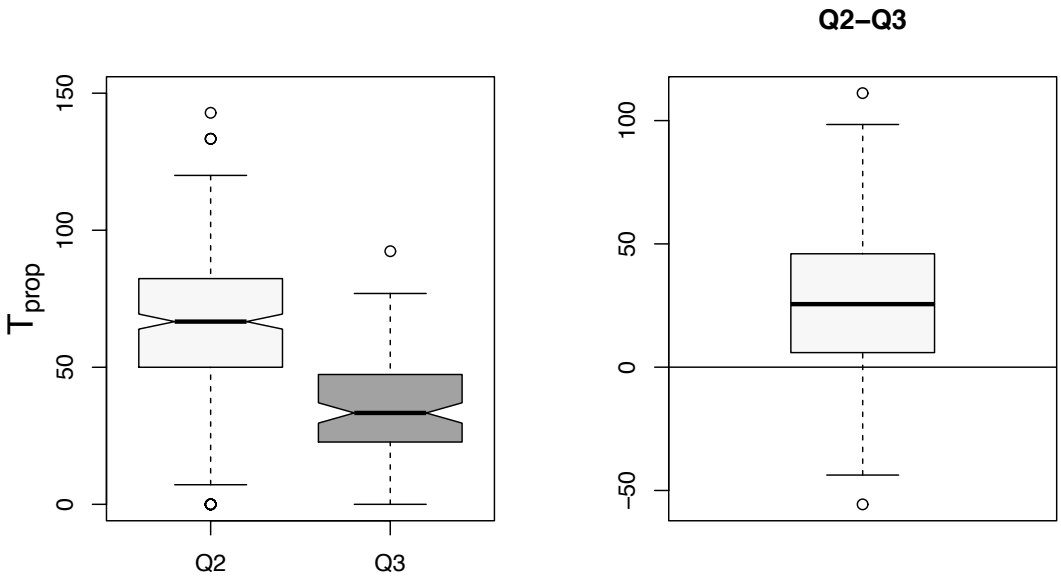

Figure 3.7. Proportional peak alignment in relation to quantity. Time of the peak (proportionally to the first stressed vowel) in relation to quantity (Q2 refers quantity two, Q3 to quantity three).

The RM-ANOVA with the dependent variable proportional peak, random variable subject and the within variable quantity[11] showed a significant effect of quantity on the proportional peak alignment ($F[1,16] = 88.4$, $p < 0.001$). Figure 3.7 shows that the peak is significantly earlier in Q3 words than in Q2 words. This finding is compatible with numerous previous studies (just to name a few: Lehiste 1960, 1979; Eek, 1983; Lippus et al., 2013) on Estonian quantity.

The following section proceeds to the analysis of intonational prominence in relation to the three fixed effects focus type (broad *vs.* narrow focus), grammatical function (subject *vs.* object) and sentence position (final *vs.* initial). First, the effect of focus type (narrow *vs.* broad) on the *peak height*, *slope* and *peak alignment* is going to be tested. Second, the effect of word order (SVO *vs.* OVS) on *peak difference* in broad and narrow focus is going to be examined.

Prior to the main analysis, Figure 3.8 presents an overview with average locations of the TPs and the TF0s annotated in the target words.

[11] ezANOVA(peakData, .(tProp), .(subj), .(quantity))

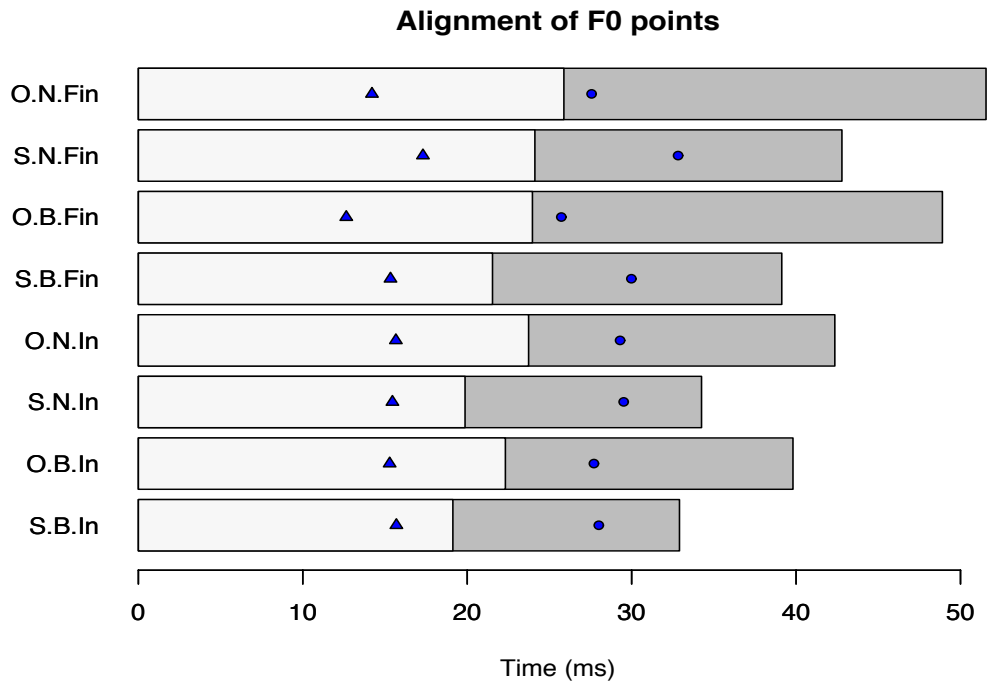

Figure 3.8. Absolute time (in milliseconds) of turning points (TP, triangles) and F0 targets (TF0, circles) in relation to the first (light grey barplots) and second syllable (dark grey barplots) in sentence-initial (In) and sentence-final (Fin) words. S refers to the subject noun phrase, O to the object noun phrase; B means broad focus contect and N narrow focus context

In Figure 3.8, it can be seen that on average TP occurred in the stressed vowel (light grey bars), whereas the TF0s appeared in the beginning of the unstressed syllable.

Figure 3.9 presents time-normalized F0 contours across the conditions. The plots at the top are for SVO word order and at the bottom for OVS. In the right-hand plots the target word was located at the beginning of the sentence; in the left-hand plots at the end of the sentence. Observe the F0 contours in black for the effects of narrow focus.

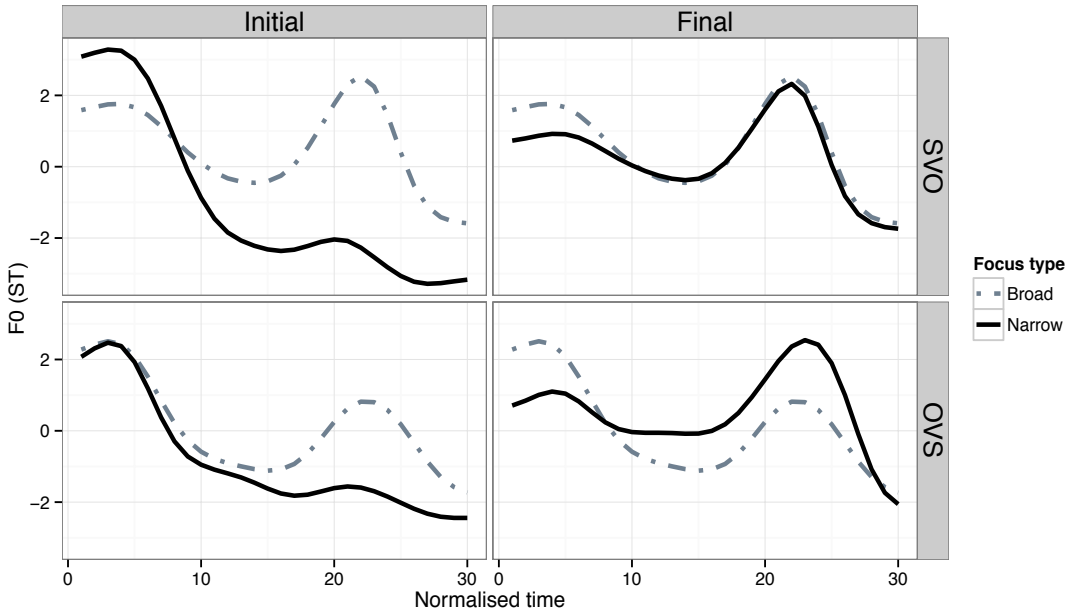

Figure 3.9. F0 contours (in semitones) aggregated over 17 speakers and 8 conditions as a function of normalized time (10 F0-values in equal time intervals for each word in a phrase). Narrow focus was either sentence-final or sentence-initial (columns), word order was either OVS or SVO (rows).

In the top-right and top-left plots, there are two prominent F0 peaks in broad focus. They appear to be almost at the same height. The black line in the top-left plot shows one prominent pitch peak on the word in focus, which indicates clear accent shift. In addition, the subject is scaled higher in narrow focus than in broad focus. In the top-right plot, the word in focus was at the end of the phrase and there appears to be no difference in scaling of narrow and broad focus. A small difference in the beginning of the phrase is observable, but its significance needs to be tested in quantified data (peak height, slope).

In the bottom-left plot, the second peak in broad focus (grey line) is considerably lower if compared to broad focus (grey line) in the plot above. It appears, thus, that word order affects F0 significantly. In the following, it is going to be tested on peak difference. Sentence-initial and sentence-final narrow focus (black lines) are similar for OVS word order and SVO word order. In both word orders, the pitch peak is aligned with the word in focus either at the beginning of the phrase or at the end of the phrase.

In the following, the contours are analyzed by the parameters of intonational prominence as defined in section 3.2.4. Figure 3.10 shows whether the peak is higher in narrow focus contexts than in broad focus context.

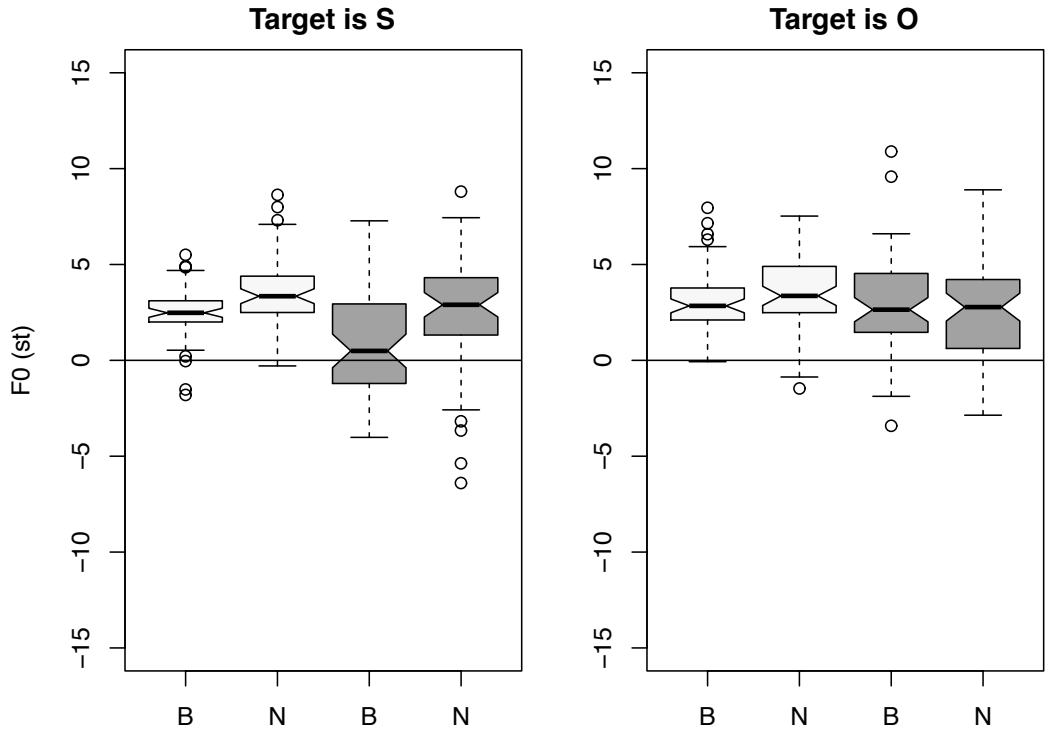

Figure 3.10. Peak height (in semitones) in relation to position in sentence (initial in light grey, final in dark grey), focus type (B refers to broad focus, N to narrow focus) and grammar (subject noun phrase (S) on the left, object noun phrase (O) on the right).

Generalized linear mixed models with the dependent variable peak height, with random variables subject and item, and with fixed factors position (final, initial), grammar (object, subject) and focus type (broad, narrow)[12] showed no significant interaction between the position, grammar and focus type. There was a significant interaction between position and grammar ($\chi^2[2] = 3.4$, $p < 0.05$) and a significant interaction between focus type and grammar ($\chi^2[1] = 14.7$, $p < 0.001$). Post-hoc Tukey comparisons showed that position had a significant effect on the subject ($p < 0.001$), but not on the object noun phrase; that grammar had a significant effect on the sentence-final position ($p < 0.05$), but not on the sentence-initial position; that focus type had a

[12] lmer(peak_st ~ focType * pos * grammar + (1|subj) + (1|target), data = peakData)

significant effect on subject ($p < 0.001$), but not on object; and finally, that grammar had a significant effect on broad focus ($p < 0.001$) but not on narrow focus.

The hypothesis that narrow focus is produced with higher peak was partly confirmed. Compatibly to the statistical evaluation, the data in Figure 3.10 shows that the peak is considerably higher on the subject in narrow focus independent of sentence position, but there is no variation in peak height on the object noun phrase.

The second acoustic parameter of intonational prominence was slope. Figure 3.11 shows whether slope is greater in narrow focus context than in broad focus context.

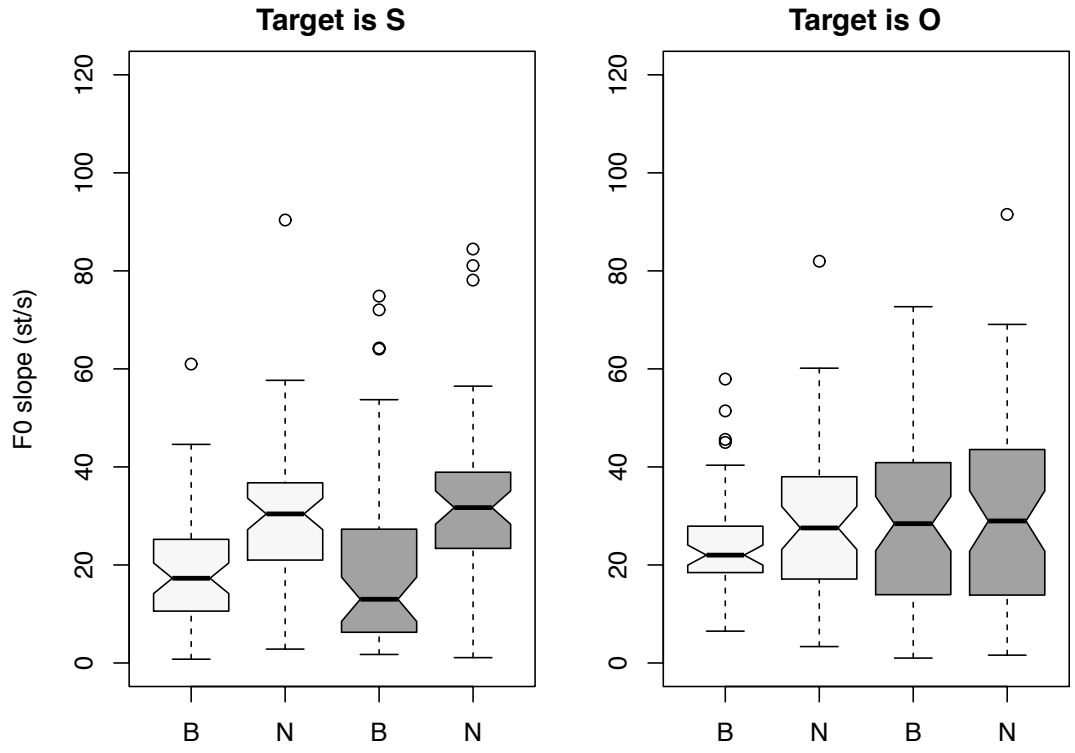

Figure 3.11. Slope (semitones in second) in relation to position in sentence (initial in light grey, final in dark grey), focus type (B refers to broad focus, N to narrow focus) and grammar (subject noun phrase (S) on the left, object noun phrase (O) on the right).

Generalized linear mixed models with the dependent variable slope and with the same fixed and random factors as before[13] showed no significant interaction between the three fixed factors. There was a significant interaction between focus type and grammar

[13] lmer(slope ~ focType * grammar * pos + (1|subj) + (1|target), data = slopeData)

($\chi^2[1] = 11$, $p < 0.001$). Post-hoc Tukey comparisons showed that focus type significantly affected the subject ($p < 0.001$), but not the object; grammar significantly affected broad focus ($p < 0.001$), but not narrow focus.

The hypothesis that the slope is greater in narrow focus than in broad focus was partly confirmed. Figure 3.11 shows that the slope was greater for narrow focus than for broad focus for subject and sentence-initial object, but not for sentence-final object. The slope for broad focus of the subject is considerably smaller than for broad focus of object in the sentence-final position.

The third acoustic parameter investigated was peak alignment. Figure 3.12 shows whether the peak is later in narrow focus context than in broad focus context.

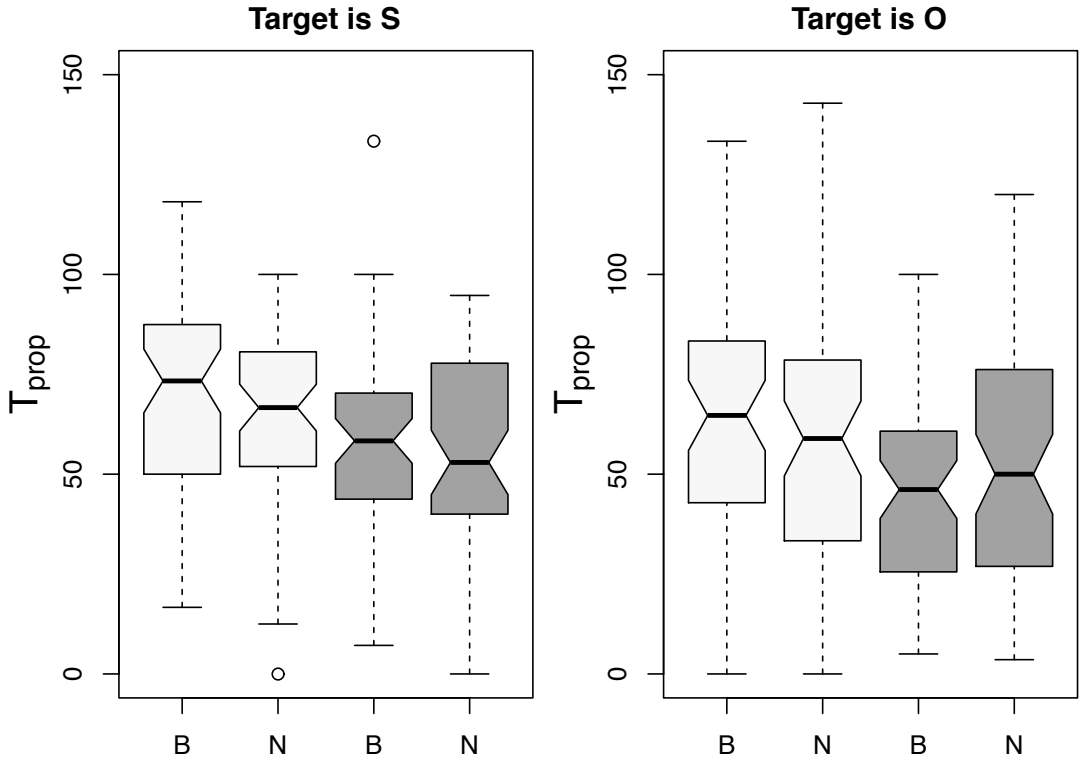

Figure 3.12. Time of the peak (proportionally to the first stressed vowel) in relation to position in the sentence (initial in light grey, final in dark grey), focus type (B means broad focus, N means narrow focus) and grammar (subject noun phrase (S) on the left, object noun phrase (O) on the right).

Generalized linear mixed models with the dependent variable peak alignment and with fixed effects position, grammar and focus type[14] showed no significant interaction

[14] lmer(peak_tProp ~ focType * pos * grammar + (1|subj) + (1|target), data = peakData)

between the position, grammar and focus type. There was a significant interaction between focus type and position ($\chi^2[1] = 4.6$, $p < 0.5$). Post-hoc Tukey tests showed that position significantly affected broad ($p < 0.001$) and narrow focus ($p < 0.05$), but there was no significant effect of focus type.

The hypothesis that the peak is later in narrow focus than in broad focus was not confirmed. Compatibly to statistical evaluation, Figure 3.12 shows that neither different focus types nor grammatical functions (subject *vs.* object) affected the alignment of the pitch peak. Interestingly, the F0 peak occurs to be earlier in sentence-final position than in sentence-initial position.

The peak alignment is affected by quantity in Estonian (Lehiste, 1960; Lippus et al., 2013), and it was also observed in Figure 3.7. Therefore, the peak alignment was additionally plotted as a function of sentence position and quantity in Figure 3.13.

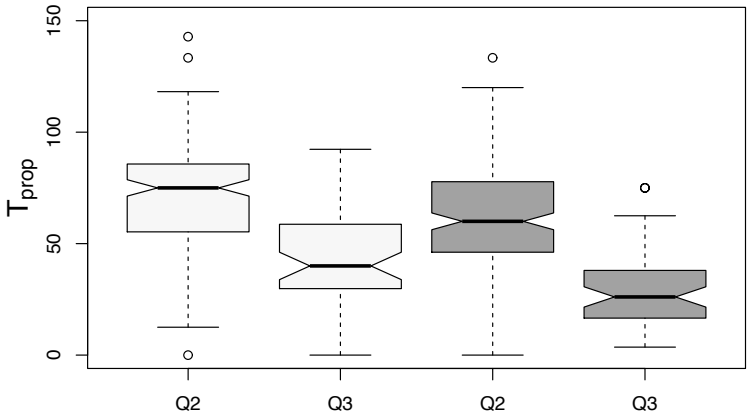

Figure 3.13. The proportional peak alignment as a function of sentence position (light grey is from sentence-initial words and dark grey comes from sentence-final words) and quantity (Q2 refers to quantity two, Q3 to quantity three).

In Figure 3.13, an important tendency in regards to the phonological status of the peak alignment can be observed: quantity-specific peak alignment is preserved despite the variation caused by sentence position.

Figure 3.14 shows boxplots with peak differences. The question whether OVS word order is possible with neutral broad focus intonation is investigated. If OVS word order is possible with broad focus intonation, then the peak difference should be similar for SVO and OVS in broad focus context.

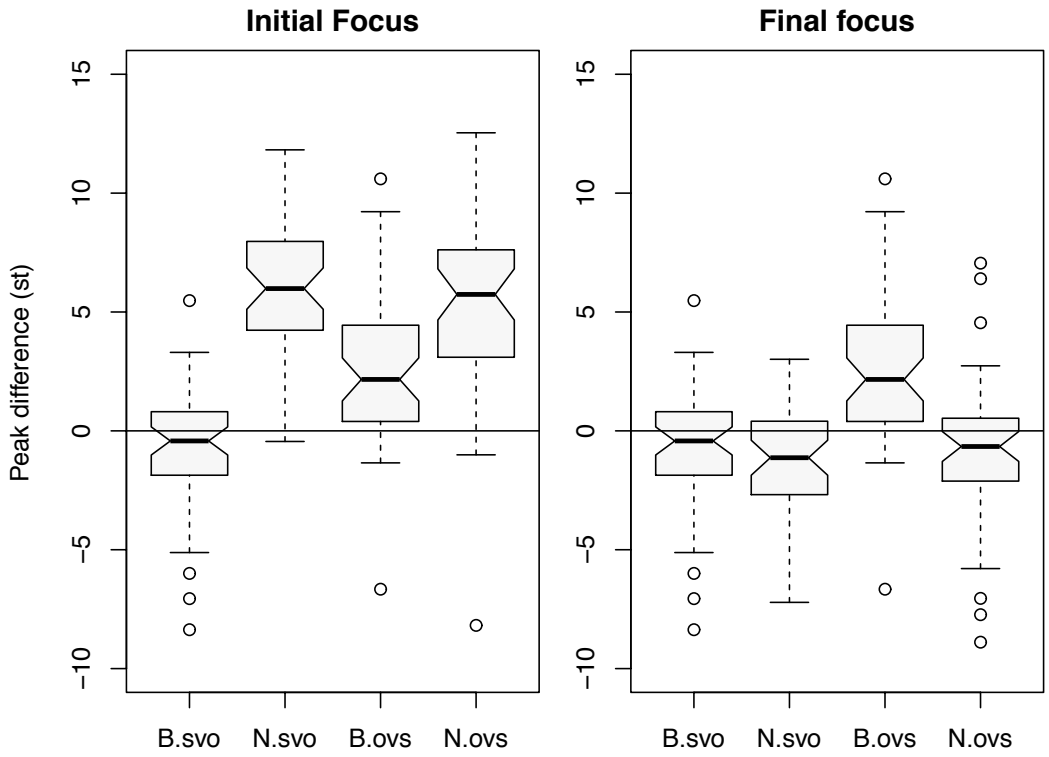

Figure 3.14. Differences between the F0 peaks (in semitone, see Figure 3.4 for illustration) in relation to word order (SVO, OVS), position of the word in focus in a sentence (initial or final) and focus type (B refers to broad focus, N to narrow focus).

Generalized linear mixed models with the dependent variable peak difference and with three two-level fixed effects position of focus (final, initial), word order (OVS, SVO) and focus type (broad, narrow) and with random effects subject and item[15] showed no significant interaction between the position, word order and focus type. There was significant interaction between focus type and word order ($\chi^2[1] = 20.7$, $p < 0.001$) and between focus type and focus position ($\chi^2[1] = 99$, $p < 0.001$). Post-hoc Tukey comparisons showed a significant effect of focus type for SVO ($p < 0.001$), but not for OVS and a significant effect of word order for broad focus ($p < 0.001$), but not for narrow focus. For the interaction between focus type and position, post-hoc Tukey tests showed a significant effect of focus type for sentence-initial and sentence-final focus ($p < 0.001$) and a significant effect of focus position for broad ($p < 0.001$) but not for narrow focus.

[15] lmer(peakDif_st ~ focType * wo * pos + (1|subj) + (1|target), data = declData)

The hypothesis that OVS word order can be produced with the broad focus intonation as measured for SVO word order was not confirmed. In Figure 3.14, it can be observed that there was almost no peak difference in broad focus of SVO sentence, whereas in broad focus of OVS sentence, it was considerably greater. This means that the first peak was higher in OVS word order than the second peak. In the narrow focus context, the position of focus significantly affected the peak difference: if the focus was located sentence-finally then the peak difference varies close to zero, which means that the two peaks in a sentence are at the same level and there is no declination of the topline; if the focus is located sentence-initially then the peak difference varies from 5 to 7 st, which means that the first peak of a phrase is considerably higher than the last one.

3.4. Discussion

The main goal of the study was to find out whether an F0 is consistently used for signalling narrow focus, and whether F0 range is scaled higher than in narrow than in broad focus. A secondary aim was to establish whether it is possible to produce an utterance that has object noun phrase in the focus position (object noun phrase of OVS word order) with a neutral broad focus intonation.

The experimental materials (see section 3.2) included target words with differing quantities (see Table 3.1). Therefore, the effect of quantity on the peak height, slope and peak alignment was tested prior to the main analysis. In the results, different quantities did not affect *peak height*. This result is consistent with a recent study by Lippus et al. (2014) who also found that the height of the F0 peak does not vary with quantities. Consistent with the results from other studies (Lehiste, 1960, 1997; Eek, 1983; Mihkla & Kalvik, 2011; Plüschke, 2013; Lippus et al., 2013), it appears that the peak alignment and the range of fall, but not the peak height, are the main characteristics of the phonetics of Estonian quantity. *F0 slope* varied with quantity, Q3 having a slightly greater F0 slope than Q2. This might indicate that there is a greater and faster F0 fall in Q3 than in Q2. *Peak alignment* was strongly affected by quantity. As predicted by Estonian lexical phonology (Lehiste, 1960; 1997; Lippus et al., 2013), the peak in Q3 occurred much earlier than the peak in Q2 word. An earlier peak and faster F0 fall indicate that F0 falls to a much lower level in Q3 than in Q2.

The first goal of the study was to investigate intonational prominence of broad and narrow focus in Estonian speech. Bearing on the results from the perception study in chapter 2 and on findings from other languages, narrow focus was expected to be intonationally more prominent than broad focus (Cooper et al., 1985; Eady & Cooper, 1986; Ladd & Morton, 1997; Swerts et al., 2002; Baumann et al., 2006; Féry & Kügler, 2008; Breen et al., 2010). Higher F0 peak, greater F0 slope, and later peak alignment was expected to signal pitch prominence. Literature suggests (Krahmer & Swerts, 2001) that intonational prominence is often perceived and produced in reference to other pitch excursions in a phrase. Therefore, peak difference was added to tonal variables.

The results showed that intonational prominence was influenced by the grammatical function of words in focus: that is, focus type had an effect on peak height and slope on the subject noun phrase in focus. In Figure 3.10 and Figure 3.11, it was observed that the peak was higher and the slope greater for subject in narrow focus. This greater prominence did not depend on the position in the sentence: subject in broad and narrow focus was produced with greater prominence sentence-initially as well as sentence-finally. However, intonational prominence did not vary with focus type (broad vs. narrow) in object noun phrases, neither sentence-initially nor sentence-finally. As was seen in Figure 3.10 and Figure 3.11, the peak and slope did not differ for narrow focus in comparison to broad focus. This was an unexpected result but actually compatible with data from other languages (Gussenhoven, 1983; Cooper et al., 1985; Gryllia, 2009; Ladd, 2008). For instance, Cooper et al. (1985) observe for American English that the F0 peak was not higher for narrow focus neither in phrase-initial nor in phrase-final words. Thus, as Cooper et al. (1985) conclude, not the local scaling of F0, but the accent shift cues focus of the sentence.

One might suspect that the peak of the narrow focus in object noun phrases was affected by quantities. As discussed above, the objects were either Q2 or Q3 words, whereas the subject noun phrases were all Q2 words (see Table 3.1). F0 plays yet an important part in distinguishing Q3 from Q2 (Lehiste 1960, 1997; Lippus, 2009; 2011; Lippus et al., 2013). There was no effect of quantity on peak height, but a slight effect on slope (see Figure 3.5 and Figure 3.6). There is a possibility that the slope of broad and narrow focus in objects might have interacted with the slope of the quantity. In the pre-analysis, the effect of focus type with Q3 words excluded was tested, but it did not change the results. Therefore, the reason why pitch excursion is not expanded in the

narrow focus of object cannot be due to quantity variation. It is possibly connected to the sentence-final position or to the grammatical properties of the object noun phrase.

The F0 expansion and therefore pitch prominence on the object noun phrase was not greater in narrow focus than broad focus which suggests that pitch accent does not unambiguously cue narrow focus on a sentence-final object. Example (3.6) adapted from Ladd (2008: 215) illustrates the acoustic situation.

(3.6)

a) I didn't give him a sandwich, I gave him [five FRANCS]$_{FOC}$.

b) I didn't give him five pounds, I gave him five [FRANCS]$_{FOC}$.

In (3.6a), the focus domain consists of two words and is therefore larger. Ladd calls it broad focus. In (3.6b), the focus domain consists of only one word and is therefore called narrow focus. Notably, the tonal representation (the pitch accented constituent is in capital letters) of different types of foci (broad *vs.* narrow) in these examples is exactly the same: in both cases, the single pitch-accented constituent is located at the end of the sentence. So, the data by Ladd demonstrate that there is no difference in pitch accent distribution between broad and narrow focus, if the constituent in narrow focus is the final object noun phrase. The phenomenon is often called as *focus projection* in syntactic accounts of sentence stress (Gussenhoven, 1983, 1999; Selkirk, 1984, 1995), which refers to the ability of the phrase-internal argument of a syntactic phrase to propagate the focus to a syntactic domain (to the whole verb phrase or the sentence) greater than the argument itself. However, this study supports an idea that the property of focus projection should be restricted only to the specific syntactic constituent and not to the phrase-final pitch accent.

The third parameter the study investigated was peak difference that scanned the shape of the F0 topline. The results indicate that the peak scaling was significantly affected by the position of the constituent in narrow focus. Figure 3.14 demonstrated that the peak difference was about 5 to 7 st when the initial word was in narrow focus. This means that the second peak in the phrase-final position (on the word in non-focus) was extremely low in pitch. Peak difference was close to or below zero when the word in narrow focus was at the end of the sentence. This suggests that there was a pitch peak on pre-focal word that is likely to have been lower than the peak on the focal word. This leads to the conclusion that the narrow focus of subject and of object were kept apart by

distribution of pitch accents in an SVO sentence: when subject was in focus, then there was only one prominent peak in a phrase; when object was in focus, then there were two peaks in a phrase. This is evidence for accent shift in Estonian focus production.

Nevertheless, broad and narrow focus of object in an SVO sentence is tonally ambiguous. The question arises whether these foci are tonally kept apart, and if they are, by which means. Gussenhoven (2007) suggests that the two types of foci are separated by the different range of F0 compression on the pre-focal constituents: in broad focus there is no compression, whereas in narrow focus there is. Peak difference in Figure 3.14 suggests this is the case, but the observation should be tested in future studies.

The second question investigated draws on the study by Vainio and Järvikivi (2007) who found that Finnish speakers compensate for the structural focus position in prosody and are able to produce information-structurally marked word order with neutral broad focus intonation (declinating pitch accents on the content words). Following Vainio and Järvikivi (2007) who investigated the sentence-final position as focus cue, it was decided to test whether there is a similar effect for the sentence-initial position in Estonian. The object in OVS word order occurs sentence-initially and has been suggested to be associated with stronger focus than the sentence-finally occurring subject (Tael, 1988; Lindström, 2006). Therefore, it was predicted that it is not possible to produce OVS word order with neutral broad focus intonation.

Figure 3.14 showed that the peak difference is close to zero in the broad focus SVO sentence, which indicates that there is almost no declination in the SVO sentence. Asu (2004) describes that the declination of the overall sentence intonation in Estonian can be described as a gradual lowering of pitch accents (topline). In the results, F0 topline did not show a downtrend. This might come about due to the relatively short sentences the speakers produced. The investigations of the declination slope (Maeda, 1976; Swerts et al., 1996; Yuan & Liberman, 2014) raise an expectation of rather steep declination slope in short phrases. Here, the three-word-sentences consisted of verbal arguments subject and object noun phrase at the end of the phrase. The proposition is that grammatical function of a sentence constituent (e.g. argument *vs.* modifier at the sentence-final position) might have influenced declination slope in addition to the sentence length. Also the reading style might have been affected the topline. Within the scope of this experiment, the conclusion is that neutral broad focus intonation of an SVO sentence is a two-peaked intonation curve without a clear downtrend in F0 topline.

Thus, the question arises whether the speakers were able to produce an OVS sentence with a similar kind of two-peaked intonation curve in the broad focus. As reported in the results, word order had an effect on peak difference in the broad focus, but not in the narrow focus. In Figure 3.14, it could be observed that the peak difference was close to zero in SVO sentence, whereas in OVS sentence it was about 2–2.5 semitones. This difference shows that the phrase-initial peak in an OVS sentence is considerably higher than the phrase-final peak. The result indicates that it may not be possible to produce an OVS sentence with a neutral broad focus intonation. This conclusion also explains why the peak and slope (Figure 3.10 and Figure 3.11) of broad and narrow focus did not differ in the sentence-initial object. If there is no possibility to produce neutral broad focus intonation with OVS word order, there is no reason to expect that the broad and narrow focus of the sentence-initial object are different from each other.

In addition, Figure 3.10 showed that the F0 peak on the sentence-final subject noun phrase was very low (close to the speaker mean (0 line)) and the slope on sentence-final subject in Figure 3.11 was the smallest in broad focus context. This data demonstrates that the F0 on sentence-final subject was severely compressed. This is further evidence that OVS word order in Estonian cannot be produced with neutral broad focus intonation.

Our data diverge from the data in Vainio and Järvikivi (2007) who showed for Finnish that the speakers are able to successfully produce a sentence, in which the sentence-final position cues focus with broad focus intonation (see the introduction in section 3.1). Vainio and Järvikivi (2007) follow from their results that it is possible to override the pragmatic implication of the sentence-final focus position by the production of the context-appropriate sentence prosody. They call the phenomenon as (prosodic) compensation of the syntactic focus position. The experiment of this study shows that it is not possible to prosodically compensate for the sentence-initial position in Estonian. The reason might be that the sentence-initial position associates to the stronger ‘emphatic’ focus and the sentence-final position to the weaker ‘neutral’ focus.

Féry and Drenhaus (2008) have shown for German, which also seems to have a sentence-initial focus position, that OVS word order is accepted only with a nuclear pitch accent on the object noun phrase. Estonian focus production shows similarly to German that word order determines the location of the strongest intonational

prominence on the object noun phrase. However, the effect of word order disappeared in narrow focus context. Estonian speakers successfully pitch-accented the sentence-final subject in OVS sentence in the context in which the subject noun phrase was in narrow focus. This result is compatible with Välimaa-Blum (1988; 1993).

Välimaa-Blum (1988) proposes for Finnish that both word orders – subject noun phrase at the beginning of the sentence as well as subject noun phrase at the end of the sentence – are neutral with respect to information structure. In her experimental study, Välimaa-Blum (1993) demonstrates that both word orders are possible with phrase-final pitch accent and with sentence-final constituent in focus. In the light of the study by Féry and Drenhaus (2008), a further study is necessary to determine whether OVS with sentence-final intonational prominence provides sufficient cues for the perception of focus on the subject noun phrase in Estonian. If it does, then German and Estonian might differ in significant aspects in terms of free word order.

As shown in Figure 3.12, the peak alignment that has been reported as the primary cue of prominence (Ladd & Morton, 1997; Kohler, 1991) was not affected by focus type in Estonian data. Peak alignment did not play any role in focus prosody and, therefore, it can be excluded from the parameters of intonational prominence. However, the data in Figure 3.12 showed that peak alignment varied together with the sentence position. The peak was later in the sentence-initial than in the sentence-final position. Figure 3.13 showed that the variation caused by sentence position does not affect the phonological peak alignment of quantities shown in Figure 3.7.

The variation of peak alignment as a function of sentence position is found to occur also in other languages. Silverman and Pierrehumbert (1990) and Prieto et al. (1995) found that peak alignment is under backward time-pressure from the upcoming prosodic boundary. An earlier peak alignment has also been related to *tonal crowding*, which means that the tonal target of pitch accent is located earlier due to the further tone, possibly due to the phrase accent (Hualde, 2002; Arvanti et al., 2006; Prieto & Torreira, 2007). At this stage of investigation, this possibility has to be excluded, while the tonal inventory in Asu (2004) does not include phrase accents for Estonian.

Plüschke (2013) also found that peak alignment in Estonian depends on the proximity of the phrase boundary: the longer the word was, the later the peak in the vowel. She concludes that the time pressure of the upcoming phrase boundary causes an earlier peak alignment in nuclear pitch accents. In the results of this study, the same

pattern occurs at the level of the whole intonational phrase. At the beginning of the utterance the peak is late because the upcoming sentence boundary is far away, whereas at the end of the utterance the peak is located early because the sentence boundary is significantly closer.

3.5. Conclusion

The results of the previous perception study raised an expectation that F0 is one of the main cues for focus in Estonian. In the current study, it was investigated using speech production whether L1-Estonian speakers systematically use pitch accent and accent shift for focus production. In addition, it was tested whether it is possible to produce OVS word order that implies a strong focus on object noun phrase with a neutral broad focus intonation. Native Estonian speakers were asked to utter sentences with different types of foci in SVO and OVS sentences, sentence-initially and sentence-finally. The recordings were analysed with respects to intonational prominence by manual annotation of F0 maxima and minima.

The results showed that speakers of Estonian apply pitch accent and accent shift for marking the focus of a sentence. However, narrow focus did not cause stronger pitch prominence on a sentence-final object noun phrase. This causes prosodic ambiguity between broad and narrow focus of a sentence-final object noun phrase that was called a prosodic effect of focus projection. The experiment demonstrated that it is not possible to produce OVS word order with neutral broad focus intonation. The data supports the hypothesis that the sentence-initial position is a stronger focus position than the sentence-final position. However, speakers produced a pitch accent on the sentence-final subject noun phrase when the context stipulated narrow focus on it. This suggests, in accordance with Välimaa-Blum (1988; 1993), that pitch accent can override structural focus position in some contexts (in case of narrow focus).

In general, the study demonstrates that pitch accent is an important cue for focus. However, the relation between focus and pitch prominence is rather complex (cf. focus projection). Sentence intonation appears to interact with word order as with sentence focus.

4. Corrective focus in Estonian[16]

Abstract

The aim of this study is to ascertain whether *narrow corrective focus* (rejection and replacement of previously mentioned information) is prosodically more prominent than narrow *new information focus* (as an answer to *wh*-question). Previous studies have shown that narrow new information focus compared to broad focus causes F0 expansion, but the acoustic difference between the two types of narrow foci (new information focus *vs.* corrective focus) is not so clear in the empirical investigations (Baumann et al., 2006; Chen & Braun, 2006; Hanssen et al., 2008; Breen et al., 2010). In chapter 3 it was shown for Estonian that pitch accent cues the location of narrow focus (initial *vs.* final) but the F0 expansion does not differentiate between broad and narrow focus. Therefore, it was tested whether greater pitch prominence is used for corrective focus in Estonian.

A speech production experiment similar to the experiment reported in the previous chapter was carried out. The participants were asked to respond to contexts that either elicited new information or corrective focus of either subject or object noun phrase. The effects of grammatical function, sentence position and focus type were investigated in connection to F0 peak, size of F0 excursion, peak alignment and vowel duration. In addition, the difference between the peaks found on sentence-initial and the sentence-final noun phrases was calculated.

The results showed that there were neither tonal nor durational effects of corrective focus in Estonian. On the basis of these results, it is concluded that the difference between correction and newness is not necessarily expressed by prosodic means, at least not in Estonian.

[16] A version of this chapter was published in the *Proceedings of the 18th International Congress of Phonetic Sciences* (Salveste et al., 2015).

4.1. Introduction

The study reported in the following chapter investigates for Estonian whether corrective focus is stronger in pitch prominence than new information focus. Production studies (Baumann et al., 2006; Breen et al., 2010) have found that there is a correlation between the shape of F0 excursion and narrow focus: the peak is higher and later and the fall steeper on the word in narrow focus than on the corresponding word in broad focus. The same kind of association has also been found in Estonian, as seen in chapter 3. An explanation is that for pragmatic reasons the word in narrow focus needs to be more salient in discourse than the word in broad focus (or in non-focus) and, therefore, F0 is assumed to be contributing to the acoustic saliency.

In pragmatics, it has been proposed that there is a number of pragmatically driven subtypes of narrow focus (see Gussenhoven, 2007). Two commonly investigated and theoretically (Halliday, 1967a; Chafe, 1976; Lambrecht, 1994) disputed subtypes are *new information focus* and *correction*. The function of the former type is to provide new information and it arises in the context of (underlying) *wh*-question. C*orrective narrow focus* rejects and replaces the incorrect information from the preceding context. See examples in (4.1) and (4.2) respectively.

(4.1)

A: Who fried an omelette yesterday?

B: DAMON fried an omelette yesterday.

(4.2)

A: Did Harry fry an omelette yesterday?

B: DAMON fried an omelette yesterday. (Breen et al., 2010)

Semantic theory of focus (Rooth, 1992) does not differentiate between the foci in examples of (4.1) and (4.2). According to Rooth (1992), any kind of expression in focus is already contrasted to the set of other possible entities that could occupy the same position in an utterance (please refer to section 1.1.1 for more explanation). Thus, according to theory, being contrastive is the inherent property of focus. The difference between the examples in (4.1) and (4.2) lies in the size of the set of alternatives: in (4.1B) the set of alternatives consists of all the entities that can figure as agents, whereas in (4.2B) the set of alternatives is made previously explicit and consists of only one

entity: *Harry*. The question is whether the size of the set of alternatives needs to have some prosodic or linguistic effect.

There are some suggestions that it does. Chafe, (1976: 36) in his account of new-given information, assumes that the contrastive or corrective focus needs to be marked prosodically and accounted separately from new-given distinction. The above-described conflict between theoretical accounts (Halliday, 1967a; Chafe, 1976; Rooth, 1992; Lambrecht, 1994) has given rise to few empirical investigations (Katz & Selkirk, 2011; Baumann et al., 2006; Breen et al., 2010 and the references there).

Researchers of intonational pragmatics (Katz & Selkirk, 2011; Baumann et al., 2006; Breen et al., 2010) have proposed that narrow corrective focus needs more emphasis than narrow new information focus. In this connection, the studies have looked for a gradual increase of F0 range in broad, narrow new information and narrow corrective focus (Baumann et al., 2006; Chen & Braun, 2006; Hanssen et al., 2008; Breen et al., 2010). As predicted, these studies found a more pronounced pitch peak in narrow new information focus than in broad focus, but the findings for F0 excursion in corrective focus were inconsistent. In American English (Breen et al., 2010), the F0 peak of a corrective focus appeared to be even lower than the peak of new information focus and in Dutch they were relatively at the same height but with a steeper fall (Hanssen et al., 2008). For standard Chinese, on the other hand, Chen and Braun (2006) found that the falling tone was indeed produced with greater F0 range in corrective than in new information focus.

In typological research, it has been proposed that different languages use different kinds of structural means for expressing the two types of foci. For example, contrast might be expressed by syntactic means and the new information focus by prosodic means (Büring, 2009; Zimmermann & Onea, 2011). In some of the Finno-Ugric languages, including Estonian, it has been shown that the sentence-initial or preverbal position is used for special emphasis – contrast, correction or exhaustive focus (É. Kiss, 1995; Vilkuna, 1998; Hovarth, 2010; Jokinen, 2005; Erelt et al., 1993). The bulk of empirical research shows, though, that the focus in those languages is also conveyed prosodically (Vainio & Järvikivi, 2006; 2007; Sahkai et al. 2013ab; chapters 2 and 3 of the thesis).

The prosodic means of focus in Estonian are not very well studied. Sahkai et al. (2013b) investigated whether there is a correlation between pitch accent type and focus

type. Their data did not show any correlation. Compatibly to the study in chapter 3, they found that there was a stronger pitch prominence in the sentence-initial word in narrow focus than in broad focus. However, they did not find difference in pitch prominence between the new information and corrective focus. The aim of the current study is to re-examine the acoustics of corrective focus in Estonian and to provide some accurate data on the theoretical conflict discussed above.

Second, the effect of word orders (SVO *vs.* OVS) is pursued further. In the previous experiment it was seen that word order played a major role in the shape of neutral broad focus intonation. OVS word order induces emphatic, possibly corrective reading on object and as shown in chapter 3, it determined the default location of the nuclear prominence at the beginning of a phrase. However, in narrow focus contexts, the effect of word order disappeared. In this chapter, the effect of word order is going to be tested for different types of focus.

4.2. Experiment

The acoustic basis for a stronger emphasis of corrective focus against the new information focus in Estonian was tested in a speech production experiment in which participants were asked to respond to questions heard over headphones regarding a series of pictures.

4.2.1. Materials

Four sentences were constructed as *target sentences*. Target sentences consisted of three two-syllable words with preferably sonorous sounds, see the list in (4.3).

(4.3) Target sentences

1. *Leena maalis vaala*. (Lena drew a whale.)
2. *Liina liimis raami*. (Liina repaired a frame.)
3. *Meeli hüüdis Loonat*. (Meeli called for Loona.)
4. *Miili kuulis Eevat*. (Miili heard Eva).

The target sentences in (4.3) were all SVO sentences that were permuted to OVS sentences, so that all the subjects and objects occurred sentence-initially as well as

sentence-finally. All the subjects were disyllabic words of quantity two. The objects in the first two sentences (*vaala*, 'whale'; *raami* 'frame') were overlong quantity (Q3); the objects in the last two sentences were long quantity (Q2).

The sentences were triggered as responses to questions or assertions called *context* (see (4.4))[17].

(4.4) Contexts

1. Somebody drew {repaired|called|heard} a whale {a frame|Loona|Eva}?
2. Leena {Liina|Meeli|Miili} drew {repaired|called|heard} something?
3. Leena drew {repaired|called|heard} poppies {a jug|Taavi|Meeri}.
4. Anna {Taavi|Reena|Aivo} drew a whale {a frame|Loona|Eva}.

The 1st and the 2nd context in (4.4) were *wh*-questions signalled by intonation. The context was meant to be interpreted so that the listener is asked to whom *somebody* and to whom or what *something* refers to. This way of presenting the *wh*-question was chosen because it elicits full-sentence-response more naturally (see section 3.2.1 for a more detailed explanation).

The 3rd and 4th context in (4.4) were meant to elicit corrective focus of the object and the subject respectively. The word *whale* (see the first target sentence in (4.3)) is meant to replace *poppies*; or *Leena* is meant to replace *Anna*. See example (4.5).

(4.5)

Context: Leena drew some poppies.

Target: (No!) Leena drew a whale!

Two word orders (SVO, OVS), two types of foci (new information *vs.* correction) either on an object or subject resulted in 8 conditions for each of the four sentences presented in (4.3). The list of the target sentences therefore consisted of 32 items.

In addition to target sentences, four other three-word-sentences called *filler sentences* were constructed. Filler sentences consisted of an object (O), a verb (V) and a sentence modifier (adverb (A)) that also varied in word order (AVO *vs.* OVA). The filler sentences (32 items) were presented with similar contexts as shown in (4.4). In total, each participant was presented with a randomized list of 64 items (32 targets and 32 fillers).

[17] See the Estonian versions of contexts in Appendix 1.3.

4.2.2. Procedure

The experiment was run exactly in the same way as described in section 3.2.2.

4.2.3. Participants

The same participants as in the experiment presented in chapter 3 took part in this experiment: ten female speakers and seven male speakers (17 altogether) between 22 and 40 years of age (mean = 28.2 years). They were either students of the University of Tartu or professionals coming from Tartu or Tallinn. They had normal or corrected-to-normal vision and hearing. The participants contributed voluntarily.

4.2.4. Analysis

Similarly to the experiment in chapter 3, the results are reported only for high pitch accent (H* or H*L). There were only a few of low pitch accents (L*) (26 instances), and for the uniformity of the slope analysis, they were excluded. Altogether 14% (77 observations) of the recorded utterances were omitted. These included utterances with low pitch accents and utterances with incorrect word order and hesitations.

The acoustic analysis of the recordings was carried out with Praat (Boersma & Weenink, 2013) similarly to analysis described in section 3.2.4. Two points in the F0 contour were manually annotated, relying on perception and visual observation of the F0 track: first, the F0 maximum or – wherever a clear F0 maximum was missing – offset of plateau, called *turning point* (TP), was annotated. Second, the end point of the conspicuous F0 fall (onset of the next plateau), called *F0 target* (TF0), was annotated. See Figure 3.2 in chapter 3 for reference. On the basis of two points TP and TF0, the data for tonal variables – F0 peak, slope, peak alignment and peak scaling were defined as in section 3.2.4.

F0 peak is either F0 maximum or F0 turning point estimated with reference to the average F0 of a speaker aggregated over all the F0 samples generated by Praat pitch analysis (default time step, pitch floor 75 Hz) (see the semitone conversion in formula (3.3) in chapter 3). If the narrow corrective focus is signalled with a stronger pitch prominence than the narrow new information, then the peak should be higher in the former type of focus.

F0 slope was defined as a main correlate of intonational prominence: the range of the F0 change was divided by the duration of the excursion. See the formula in (3.4). The more expanded the F0 excursion is, the greater slope is expected. The F0 slope is expected to be greater in narrow corrective focus than in narrow new information focus.

Peak alignment was defined as the time from the vowel onset to the time of the F0 peak in proportion to the duration of the first stressed vowel, see the formula in (3.5). In the study in chapter 3, the peak alignment was not affected by narrow focus. This study explores whether it is affected by corrective focus instead. Based on the results of earlier studies on other languages (Kohler, 1991; Ladd and Morton, 1997), the peak is expected to be later in narrow corrective focus than in narrow new information focus, if narrow corrective focus is signalled by a stronger degree of intonational prominence.

For *peak difference*, the difference between the F0 peaks was calculated by subtracting F0 (in semitones) of the last peak from the first peak in an utterance, see Figure 3.4 for reference. As discussed in chapter 3, peak difference scans the shape of the F0 topline. The results presented in the previous chapter showed that word order (SVO *vs.* OVS) affected peak difference in broad focus context, but not in narrow focus context. In this experiment the interaction between word order and two types of foci is investigated.

Previous studies (Baumann et al., 2006; Breen et al., 2010, Sahkai et al., 2013a) have shown that the duration rather than F0 expansion signals corrective focus, therefore, the duration of the first stressed vowel was added to the analysis. The duration of the first stressed nucleus was chosen, because the duration of the long vowel in nucleus is not expected to vary with grammatical function of the word, whereas the duration of the word is. As can be seen in the materials in (4.3), the object noun phrase had the partitive ending *-t* at the end of the word, whereas subject did not. Vowels of the stressed syllable were both long. It is predicted that the stressed vowel is longer in narrow corrective focus than in narrow new information focus.

The peak height, slope, peak alignment, difference and duration of the first vowel are defined as parameters of intonational prominence. The goal is to investigate whether the dependent variable increases in the corrective focus in comparison to the new information focus. In addition, the effect of the interaction between word order (SVO *vs.* OVS) and corrective focus is investigated for peak difference (global peak

scaling. All the statistical tests were carried out with software R for statistical computing ("R Development Core Team", 2014).

4.3. Results

Figure 4.1 first gives an overview of the average location of TPs and TF0s in the noun phrases.

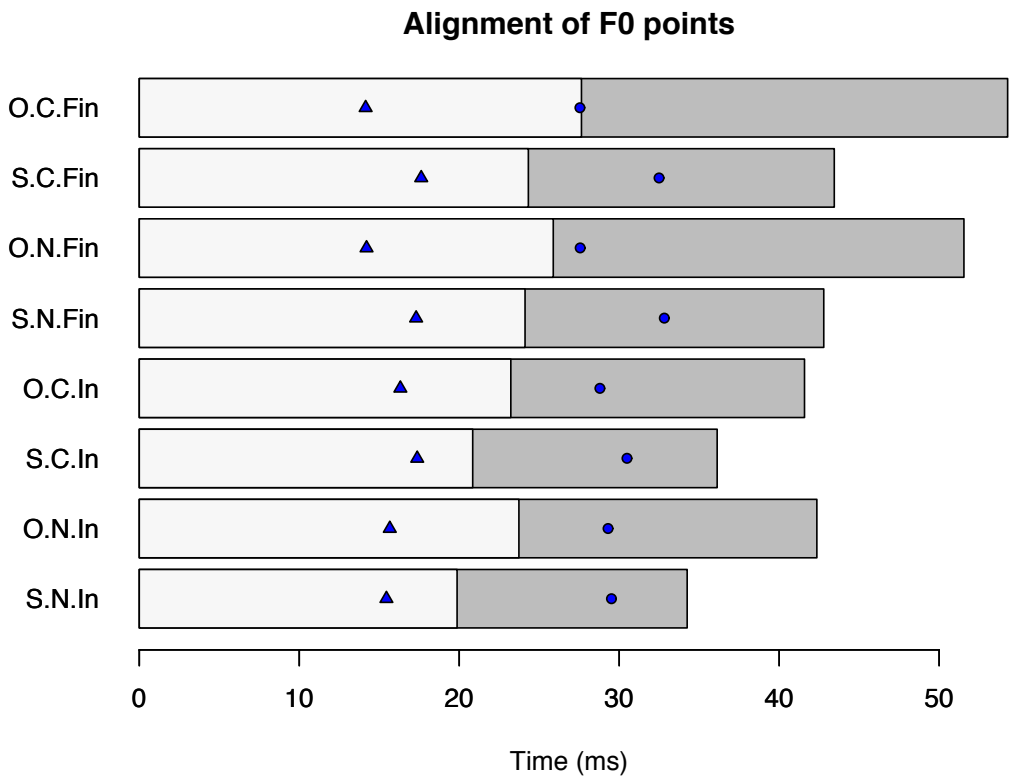

Figure 4.1. Absolute time (in milliseconds) of turning points (TP, triangulars) and F0 targets (TF0, circles) in relation to the first (light grey barplots) and second syllable (dark grey barplots) in sentence-initial (In) and sentence-final (Fin) position. S refers to subject noun phrase, O to object noun phrase, N means new information focus, C means corrective focus.

Most of the TPs occurred in the stressed vowel (light grey bars), whereas the TF0s appeared at the beginning of the unstressed syllable. The mean duration between the points (duration of the F0 excursion) was 138 ms and the standard deviation 4 ms.

In the following, the five different dependent parameters (peak height, slope, peak alignment, peak difference and duration of the first stressed vowel) were tested in relation to the three fixed factors:

1) focus type: narrow new information focus (N) *vs.* narrow corrective focus (C)
2) grammatical function subject (S) *vs.* object (O)
3) sentence position (final *vs.* initial).

Figure 4.2 shows peak height.

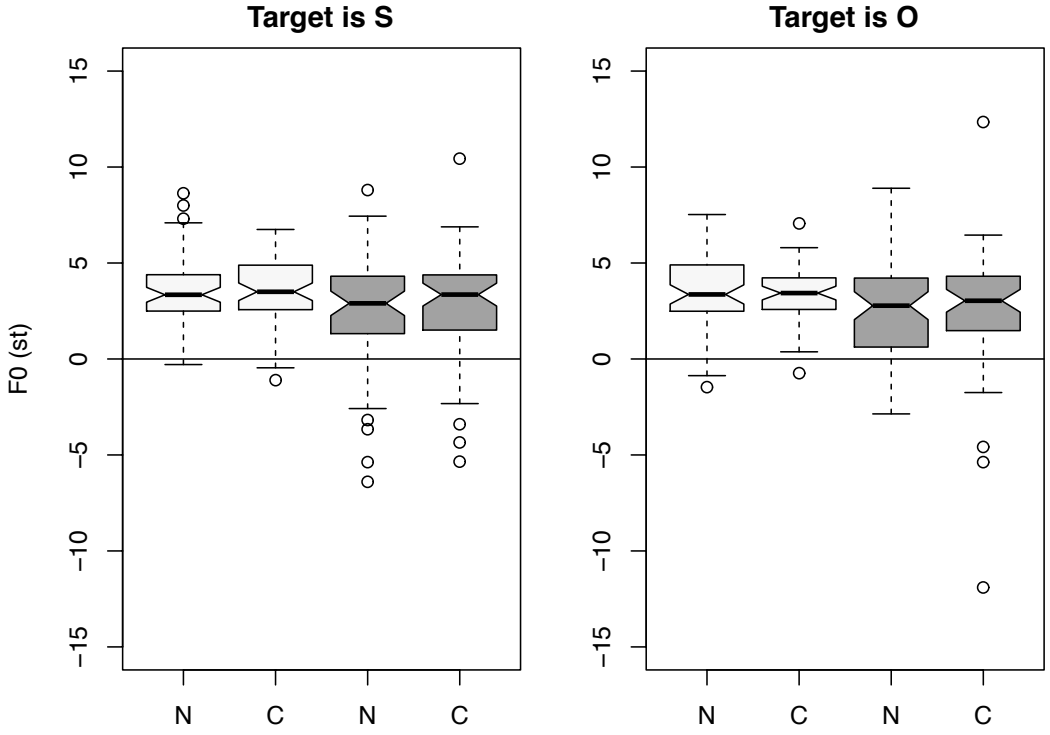

Figure 4.2. Peak height (in semitones) in relation to position in a sentence (initial in light grey, final in dark grey), and as a function of focus type (N refers to new information, C refers to correction) and grammar (subject (S) on the left, object (O) on the right).

Generalized linear mixed models with the dependent variable peak height, random effects subject and item and fixed effects position (initial *vs.* final), focus type (new information *vs.* correction) and grammatical function (subject *vs.* object)[18] showed no significant interactions between the factors and no significant main effects.

The hypothesis that the peak is higher in narrow corrective focus than narrow new information focus was not confirmed. As can be seen in Figure 4.2, peak height

[18] lmer(peak_st~grammar * is * pos + (1|subj) + (1|target), data = peakData)

varies between 2 and 4 semitones and there were no effects of focus type, grammar or position. Although, it can be observed that sentence-final peaks tend to be slightly lower than sentence-initial peaks.

The second acoustic parameter of intonational prominence – the slope – estimates the steepness of the F0 excursion. Figure 4.3 plots the slopes as a function of focus type, sentence position and grammatical function.

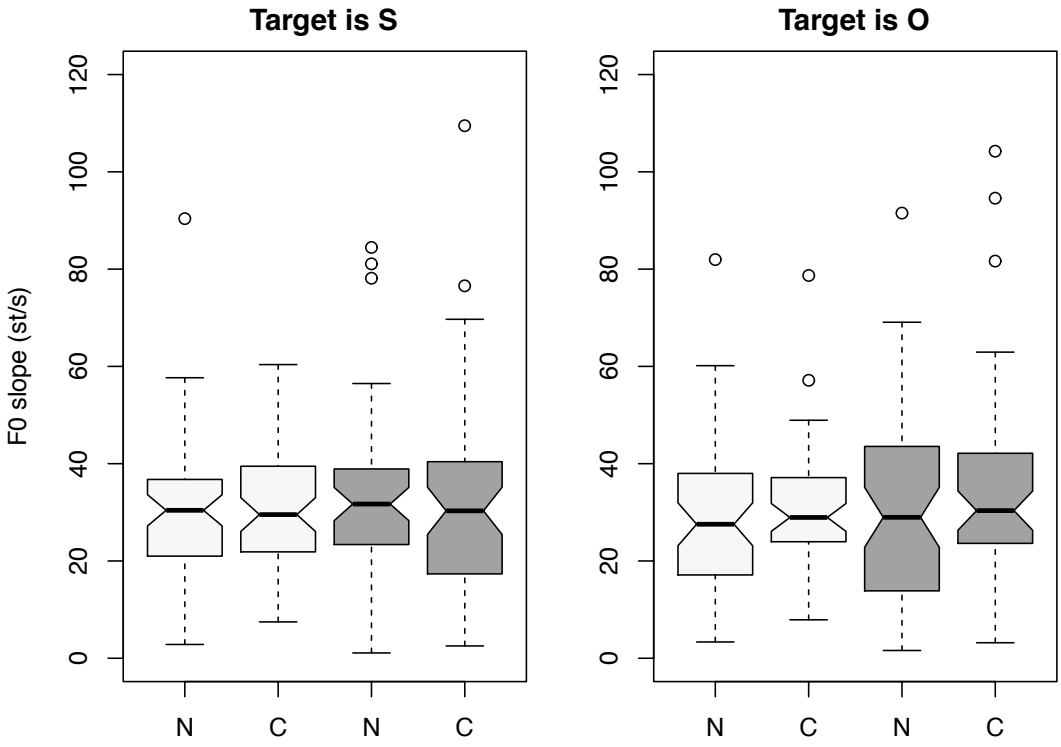

Figure 4.3. Slope (semitones per second) in relation to position in a sentence (initial in light grey, final in dark grey), to focus type (N refers to new information, C refers to correction) and to grammar (subject noun phrase (S) on the left, object noun phrase (O) on the right).

Generalized linear mixed models with the dependent variable peak height, with random effects subject and item and with fixed effects position (initial *vs.* final), focus type (new information *vs.* correction) and grammatical function (subject *vs.* object)[19] revealed no significant interactions between the factors and no significant main effects.

The hypothesis that the slope is greater in narrow corrective focus than in narrow new information focus was not confirmed. In Figure 4.3, it can be seen that all

[19] lmer(slope~grammar * is * pos + (1|subj) + (1|target), data = slopeData)

the boxplots are at the same size: the slope varies between 20 and 40 st/s and the explanatory effects do not influence the size of the slope.

Figure 4.4 plots peak alignment as a function of grammar, position and focus type. The question examined is whether the peak is aligned later for narrow corrective focus than for new information focus.

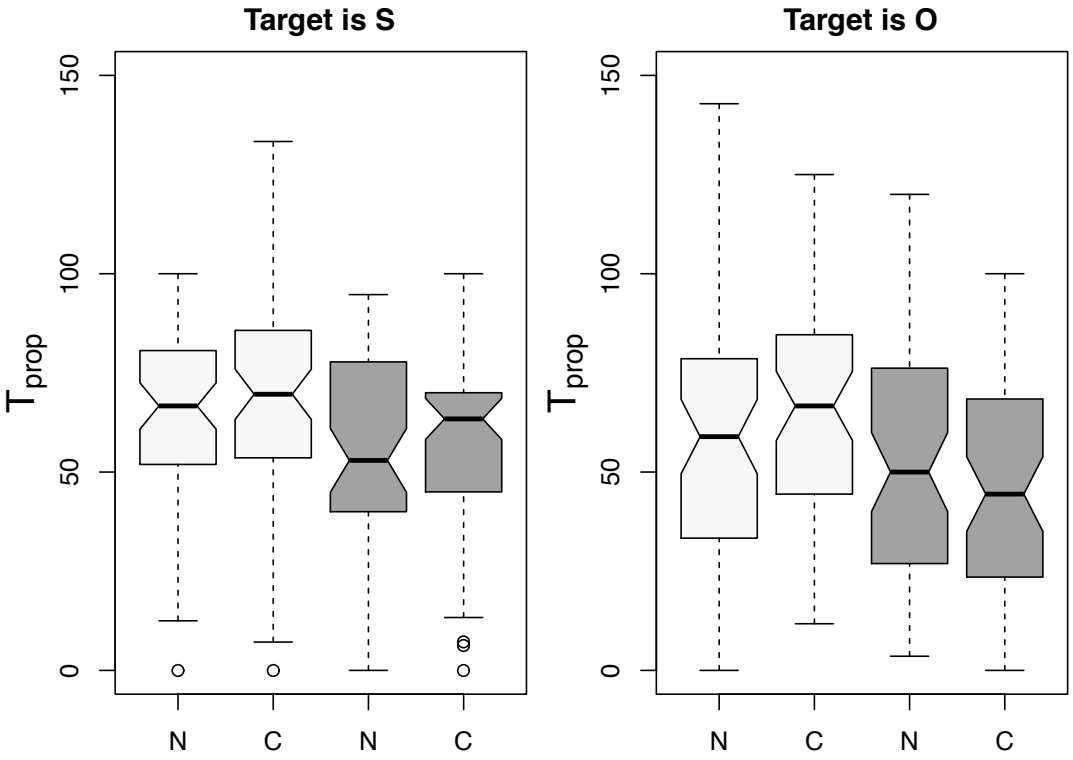

Figure 4.4. Time of the peak (proportionally to the first stressed vowel) in relation to position in a sentence (initial in light grey, final in dark grey), to focus type (N refers to new information, C to correction) and to grammar (subject noun phrase (S) on the left, object noun phrase (O) on the right).

Generalized linear mixed models with the same fixed and random effects as before and with the dependent variable peak alignment[20] showed no significant interactions between the factors, but a significant main effect for position ($\chi^2[1] = 30$, $p < 0.001$) occurred.

The hypothesis that the peak is later in narrow corrective focus than in narrow new information focus was not confirmed. However, Figure 4.4 shows the effect of

[20] lmer(peak_tProp~grammar * is * pos + (1|subj) + (1|target), data = peakData)

sentence position: the peak is located earlier in sentence-final words than in sentence-initial words.

Figure 4.5 demonstrates the average location of the peak (vertical blue line) within the stressed vowel (vertical grey and black lines) of the word in focus on the background of the pitch contours. The contours were not plotted in relation to the factors of word order and of focus type, since they did not turn out to be significant (a more detailed Figure in relation to all three factors can be seen in Appendix B.3).

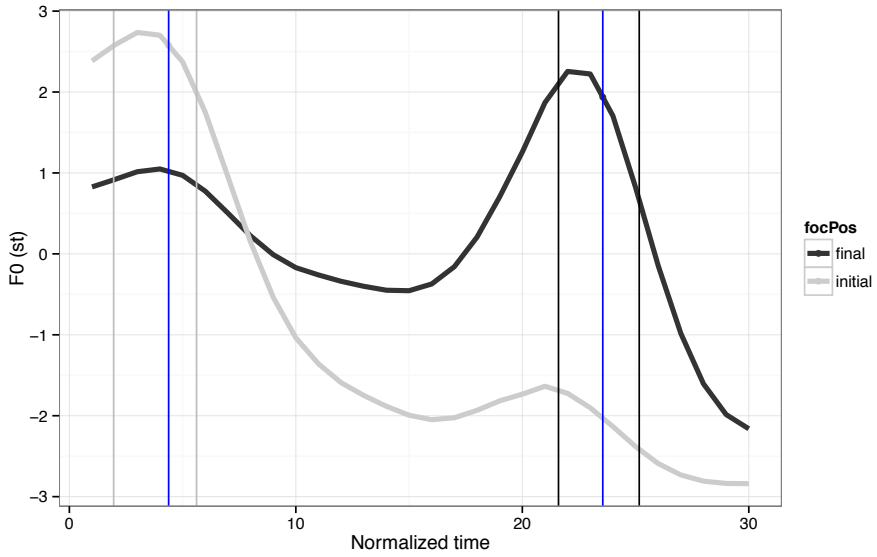

Figure 4.5. Aggregated F0 over utterances as a function of normalized time (10 F0-values in equal time-intervals for each word in a phrase). *Black contour* is the average F0 from the focus-final and the *grey contour* from the focus-initial utterances. *Blue vertical line* is the location of the peak, *grey vertical lines* the boundaries of stressed vowel in a sentence-initial word in focus and *black vertical lines* the boundaries of stressed vowel in a sentence-final word in focus.

Figure 4.5 demonstrates that in sentence-initial position the peak is located closer to the end of the vowel than in sentence-final position. It is approximately in the middle of the vowel and further away from the offset. As discussed in chapter 3, the effect of the later peak in sentence-final constituent might arise due to the left-hand intonational phrase boundary.

In the evaluation of peak difference, the effect of grammar was replaced by the effect of word order (SVO *vs.* OVS). The effect of position reflects the position of the word in focus in the sentence: combination of initial focus and SVO means that the

subject of SVO word order was in focus. Peak differences close to zero indicate that phrase-initial and phrase-final peaks were at the same level, a difference greater than zero means that the phrase-initial peak was higher than the phrase-final peak, and a difference below zero that the phrase-initial peak was lower than the phrase-final peak. Figure 4.6 plots peak difference as a function of focus type, position of focus and word order.

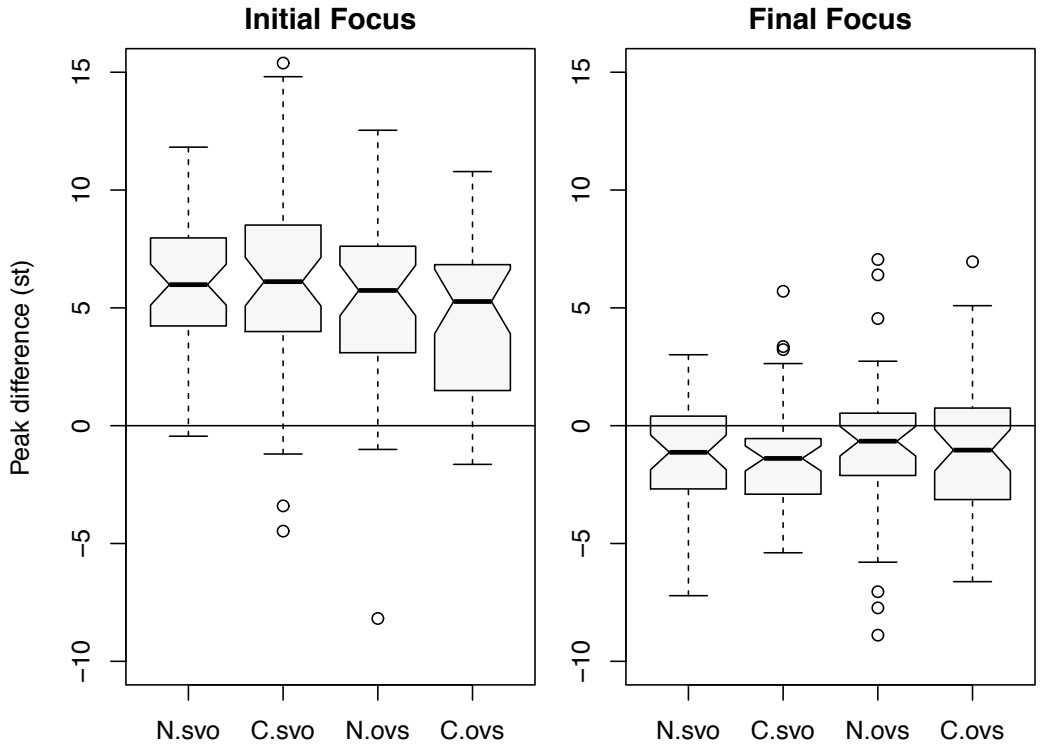

Figure 4.6. Differences between the F0 peaks (in semitone) in relation to word order (SVO, OVS), to position of the word in focus in a sentence (initial or final) and to focus type (N referes to new information focus, C to corrective focus).

Generalized linear mixed models with the dependent variable peak difference, with random effects subject and item and with fixed effects focus position (initial *vs.* final), focus type (new information *vs.* correction) and word order (SVO *vs.* OVS)[21] showed a significant interaction between word order and position ($\chi^2[1] = 7$, $p < 0.01$). Post-hoc Tukey tests showed that position affected the peak difference in SVO ($p < 0.001$) and

[21] lmer(decl_H_st ~ focType * wo * pos + (1|subj) + (1|target), data = declData)

OVS word order ($p < 0.001$); word order affected the sentence-initial position ($p < 0.5$), but not the sentence-final position.

The model was consistent with the result seen in the Figure 4.6: corrective focus did not interact with word order. For sentence-initial focus, the peak difference was smaller in OVS than in SVO word order independent of focus type. The peak difference was significantly affected by the position of focus: the difference varied between 4 and 8 st if the word in focus was in sentence-initial position, but between –3 and 0 if the word in focus was sentence-final.

The results above unanimously demonstrated that the pitch does not differentiate between narrow new information and narrow corrective focus. Therefore, additional phonetic parameter was investigated. Figure 4.7 provides data for the question whether the two focus types differ in duration of the stressed vowel.

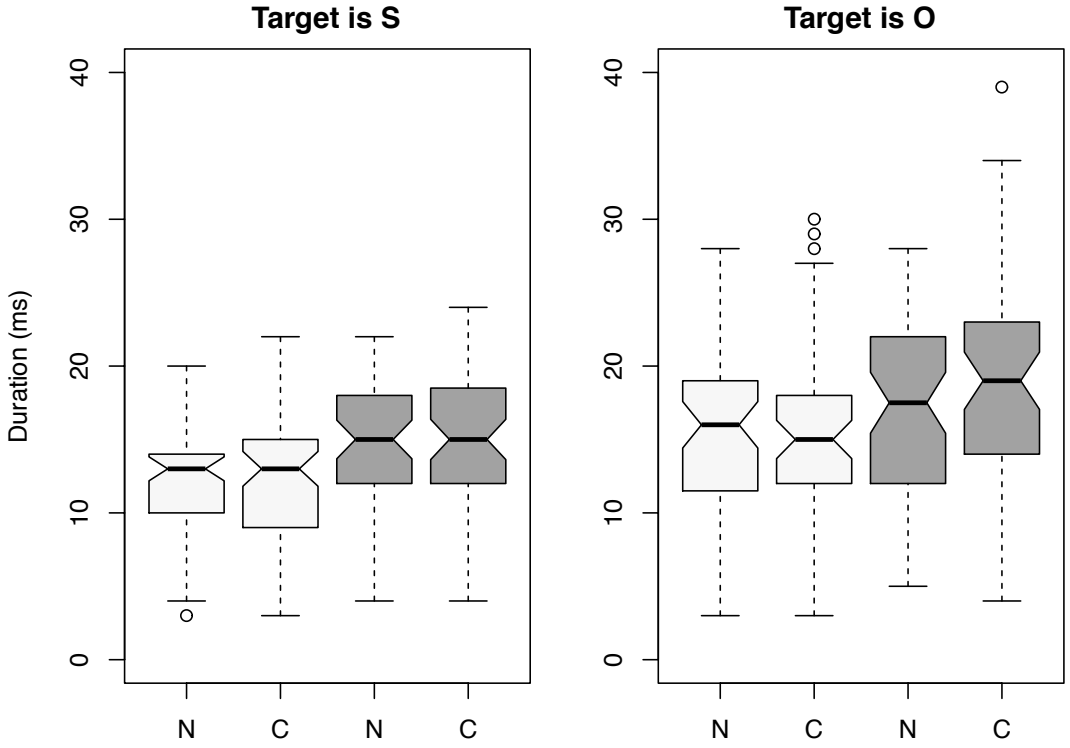

Figure 4.7. Vowel durations (in milliseconds) in relation to position in a sentence (initial in light grey, final in dark grey), to focus type (N refers to new information, C to correction) and to grammar (subject noun phrase (S) on the left, object noun phrase (O) on the right).

Generalized linear mixed models with the dependent variable vowel duration, with random effects subject and item and with fixed effects position (initial *vs.* final), focus type (new information *vs.* correction) and grammatical function (subject *vs.* object)[22] showed a significant main effect of position ($\chi^2[1] = 45$, $p < 0.001$).

The hypothesis that the stressed vowel in narrow corrective focus is longer than in narrow new information focus was not confirmed. For the significant effect of position, Figure 4.7 shows that a vowel in the sentence-final position was longer than in the sentence-initial position independently of grammar or focus type. This can be attributed to the phrase-final lengthening.

4.4. Discussion

The experiment was run to test whether narrow corrective focus causes greater pitch prominence on the constituent in focus, if compared to narrow new information focus. In addition, the effect of focus type (narrow new information *vs*. narrow corrective focus) was investigated in relation to word order (SVO *vs*. OVS). Peak height, slope, peak alignment, peak difference and duration of the first vowel were defined as parameters of prosodic prominence.

There was no effect of focus type on vowel duration. Duration was affected by position: sentence-final vowels were longer than the ones in sentence-initial position, possibly as a result of phrase-final lengthening (cf. Plüschke, 2013). There was again no effect of focus type on peak alignment, which is in line with the findings described in chapter 3, where it was argued that the timing of the peak depends first on the word quantity and second on the upcoming phrase boundary. As for the F0 expansion, the results showed that focus type affected neither peak height nor F0 slope: this finding therefore does not support the prediction of phonetic difference between *narrow new information* and *narrow corrective* focus.

Hence, narrow corrective focus is not signalled by a stronger pitch prominence than narrow new information focus. The result is consistent with Sahkai et al. (2013b). As discussed in the introduction, some studies have shown that corrective focus is associated with a lower F0 peak (Breen et al., 2010) or steeper F0 fall (Hanssen et al.,

[22] lmer(vDur ~ focType * grammar * pos + (1|subj) + (1|target), data = slopeData)

2008). The results of the present study show by contrast that there is no difference in pitch prominence between the two types of narrow focus, at least on the word that is in focus.

However, different focus types may possibly be signalled in the pre-focal prosody (see also Gussenhoven, 2007). As was discussed above, peak difference scans the F0 topline of an utterance and might therefore capture a stronger pre-focal F0 compression. For stronger F0 compression, peak difference is expected to be greater in corrective than in new information focus. The results showed that focus type did not influence peak difference. Thus, this is tentative evidence that, at least in Estonian, correction is not signalled by prenuclear or postnuclear intonation. However, this needs further investigation.

A sentence position of the word in focus affected peak difference. Peak difference was between –3 and 0 st when the word in focus was at the end of the sentence, whereas it was between 4 and 8 st when the word in focus was at the beginning of the sentence. This can be most likely explained by the presence of a prenuclear accent on the first word in the case of utterance-final focus. The first word was mentioned in the previous context and, therefore, represented given information. This result indicates, thus, that the given information preceding the focus carries a prenuclear pitch accent.

The result of the experiment can be explained within the framework of the alternative semantics of focus (Rooth, 1992). This account does not distinguish between the semantics of the two types of focus: both the novelty of the word and a rejection of the previously mentioned word are inherently contrastive. As discussed in the introduction, the difference between the two foci lies in the size and the characteristics of the set of alternatives: in the corrective focus it consists of a single explicit alternative, whereas in new information focus it consists of an uncountable set of entities that can take the same semantic role. The result of the experiment is consistent with the view that the size and the characteristics of the set of alternatives are not signalled by prosody.

As discussed above, it is theoretically known for Estonian that the object noun phrase at the beginning of a sentence is in ‘emphatic’ focus (Erelt et al., 1993; Lindström, 2006). Word order with implicit ‘emphatic focus’ was included in the experimental design. There was a slight tonal effect on sentence-initial object noun

phrase. Figure 4.6 indicated that word order had an effect on peak difference and that the first peak was lower on object noun phrase for OVS word order. This result might indicate that an object in initial position is in corrective focus simply by word order. Since an object in initial position is such a strong marker of focus, there is no need for the speaker to provide additional prosodic cues to signal it as being in special type of focus. This result is also somehow parallel to the result reported in the second experiment of Breen et al., (2010). Contrary to their prediction they find that the F0 peak in narrow corrective focus is scaled lower than in narrow new information focus. This curious effect indicates that the tonal cue for corrective focus is rather a lower than a higher peak. If this would be true, then the phrase-initial lower F0 and the focus position of sentence-initial object would be consistent with each other.

In the experiment reported in chapter 3, it was seen that word order had an influence on intonation in broad focus but not in narrow focus context. An OVS word order caused the nuclear prominence to be shifted to the beginning of the phrase in the broad focus context. Therefore, OVS word order induces an emphatic, possibly corrective reading on the object noun phrase, which is reflected in the speakers' preference to shift nuclear pitch accent to the beginning of the phrase. The result suggested that word order shapes sentence-intonation in such a way that there is a nuclear pitch accent on the object if a sentence has an OVS word order. However, when a sentence with OVS word order was embedded into a context with narrow focus – no matter whether with focus on the subject or on the object noun-phrase, nuclear pitch accent was located on the constituent in focus. This was also seen in the current study. Figures 4.2, 4.3, and 4.6 showed that there was no effect of grammar or word order on sentence intonation. The results reported in this chapter confirm the observation in chapter 3. Hence, accent shift is possible in a sentence with SVO as well as OVS word order.

4.5. Conclusion

A speech production experiment was carried out in order to investigate intonational prominence of different focus types in connection to sentence position and grammatical function in Estonian. Based on previous studies (Krahmer & Swerts, 2002; Baumann et al. 2006; Breen et al., 2010), F0 peak height, F0 slope, peak alignment, peak difference

and vowel duration was chosen as parameters for detecting the degree of intonational prominence. The two types of narrow focus were expected to differ in peak height, F0 slope, peak alignment, in duration and in peak difference. The study in chapter 3 showed that the effect of word order disappeared in the context of narrow focus. Therefore, it was proposed that the word order OVS is sensitive to different pragmatic types of focus, and the interaction between the narrow corrective focus and word order was investigated against the interaction between word order and new information focus.

The results showed that there is no prosodic difference between the two types of narrow foci (new information focus *vs.* corrective focus). Firstly, the findings indicate at least for Estonian that narrow corrective focus does not cause significantly stronger pitch prominence than narrow new information focus. However, further investigation is needed to ascertain whether prenuclear prosody cues focus type. The two focus types might still be perceptually distinguished. The question remains what makes the corrective focus more 'emphatic': its acoustics or its semantics? Secondly, if embedded into narrow focus context, there was no influence of word order (SVO *vs.* OVS) on sentence prosody. In other words, the pragmatic implications provided by word order could be overridden by the pragmatic implications contained in the previous context. Possibly, pragmatic functions other than investigated in the study may account for OVS word order in Estonian.

5. Phonetics of givenness in Estonian

Abstract

In this study the question whether *givenness* (as a term introduced by Chafe, 1976; Prince, 1981; 1992; Baumann, 2006) causes F0 compression in Estonian, and whether the predicted F0 compression is affected by the grammatical function or by the position of the expression in a sentence. Investigation of intonation languages has shown that F0 on given expressions is often severely compressed (Terken & Hirschberg, 1994; Xu & Xu, 2005; Baumann, 2006; Baumann & Riester, 2012), a phenomenon referred to as *deaccentuation*. Deaccentuation occurs in English and German but not necessarily in other languages (Cruttenden, 2006). Terken and Hirschberg (1994) suggest that even for English, deaccentuation might interact with position in a sentence and with grammatical function. The studies in previous two chapters have indicated that, in Estonian, there is a 'non-prominent' pitch accent in the pre-focus position. Based on hypotheses in Terken and Hirschberg (1994) and the reports in chapter 3 and 4, the current study investigates the interaction between deaccentuation, sentence position and grammatical function.

An experiment with a speech elicitation task was run. The participants were asked to utter sentences with embedded target words that varied in their information status (new *vs.* given), grammatical function (subject *vs.* object) and sentence position (initial *vs.* final). In order to estimate the effect of givenness, range of F0 excursion and vowel duration were analysed. In addition, declination of pitch peaks was investigated as a measure of peak difference.

The results showed that givenness was affected by sentence position: given information that occurred sentence-initially before the focus carried a prenuclear pitch accent, whereas given information at the end of the sentence was deaccented. There was no effect of grammar. Importantly, the results showed that similarly to intonation languages, givenness is deaccented in Estonian. In addition, the pre-focal and post-focal position affects the phonetics of givenness. Hence, there is no direct correlation between range of F0 excursion and information structural value of a sentence constituent.

5.1. Introduction

The aim of the present study is to investigate deaccentuation of contextually given information in Estonian. The first question is whether F0 is compressed to such a degree that it can be considered as deaccentuation. The second question builds on the positive result of the first question: if it occurs that givenness causes deaccentuation, does the degree of F0 compression depend on the position in relation to focus (pre-focal *vs.* post-focal) or grammar (subject noun phrase *vs.* object noun phrase) of the expression.

Givenness is an information-structural term that refers to information that is known to all speech participants in a particular discourse (Chafe, 1976; Lambrecht, 1994; Baumann, 2006). Given or shared information is assumed to consist of referents that are activated and identifiable (to different degrees) to the speech participants (Chafe, 1976; Lambrecht, 1994; Baumann, 2006). The studies have shown that F0 in expressions containing activated referents is severely compressed – *deaccented* (Brown, 1983; Terken & Hirschberg, 1994; Xu & Xu, 2005 for English, Baumann, 2006; Baumann & Riester, 2012 for German; Swerts et al., 2002 for Dutch). This mechanism is proposed to guarantee that a constituent in focus is phonetically highlighted and easily recognizable (e.g. Krahmer & Swerts, 2001; Xu & Xu, 2005, numerous studies on word and phoneme recognition). Further perception studies have shown that compressed intonation or a missing accent significantly contributes to the recognition of a contextually or lexically activated word in a sentence (Baumann & Hadelich, 2003; Baumann & Grice, 2006).

However, the occurrence of deaccentuation is not defined only by contextual properties, but depends on properties of a particular language (see Cruttenden, 2006). The question whether a language deaccents on given expressions is therefore not a trivial one. Eva Gårding (1981) has observed for Swedish that given information carries prominent F0 excursions even in the post-focal position. Gårding (1981: 152) suggests that prominent F0 excursions found on the given words are lexical pitch accents that need to be preserved for morpho-lexical meanings. However, not only a pitch accent language (like Swedish) resists F0 compression, but also intonation languages do that. Krahmer and Swerts (2001) have shown that in comparison to Dutch speakers, the speakers of Italian have difficulties in deaccenting the previously mentioned nouns.

Similarly, Alan Cruttenden (2006) has reported that pitch accenting on given information is quite common to all Romance languages.

Some studies (e.g. Terken & Hirschberg, 1994) have suggested that rhythmical reasons might also interfere with expected deaccentuation in pre-focal position. Jacques Terken and Julia Hirschberg (1994: 127) have suggested that the information status "is not sufficient to predict speaker's decisions to accent or deaccent expressions in discourse". Terken and Hirschberg (1994) have given rise to the hypothesis that givenness in sentence-initial position is not signalled by the absence of pitch accent but rather by varying pitch range of the pitch accent. This means, that a previously mentioned referring expression is likely to carry (non-prominent) prenuclear pitch accent. Consistently, Caroline Féry and Frank Kügler (2008) suggest for German that givenness in the prenuclear position cannot be signalled by the absence of pitch accent but rather by just compressing the pitch range of the pitch accent.

The current study aims to investigate the phonetics of given information in Estonian, which might offer an interesting test-bed for the interaction between word- and sentence-level prosody. At the morpho-lexical level, Estonian has a three-way quantity distinction that is signalled largely by a tonal cue (see section 1.1.4; Lehiste, 1960; Eek, 1983; Lippus et al., 2009; 2010). This word-level tonal characteristic makes it more similar to pitch accent language such as Swedish. Studies on sentence intonation show, however, that focus is highlighted by pitch accent (see e.g. chapter 3,4, Sahkai et al. 2013a). In this sense, Estonian is similar to intonation languages. But to which type of intonation language: Romanic type of language that resists deaccentuation or Germanic type that deaccents given information? Interestingly, a corpus study on Estonian (Lippus et al., 2013) provides evidence that the tonal cues of quantities are not preserved in non-focus position. The non-focus words are, thus, expected to be deaccented.

The experiments in chapters 3 and 4 showed that the sentence-initial focus is highlighted by accent shift, as illustrated in Figure 5.1. The accent shift involves deaccentuation of non-focal words. The peak difference indicated that there was a (non-prominent) pitch accent in the pre-focus position, but a serious F0 compression in the post-focus position. Therefore, the experiment of this study tests whether and to what extent contextually given words are deaccented in pre- and post-focal position. Post-

focal words are predicted to be deaccented, whereas pre-focal expressions are expected to carry a prenuclear pitch accent.

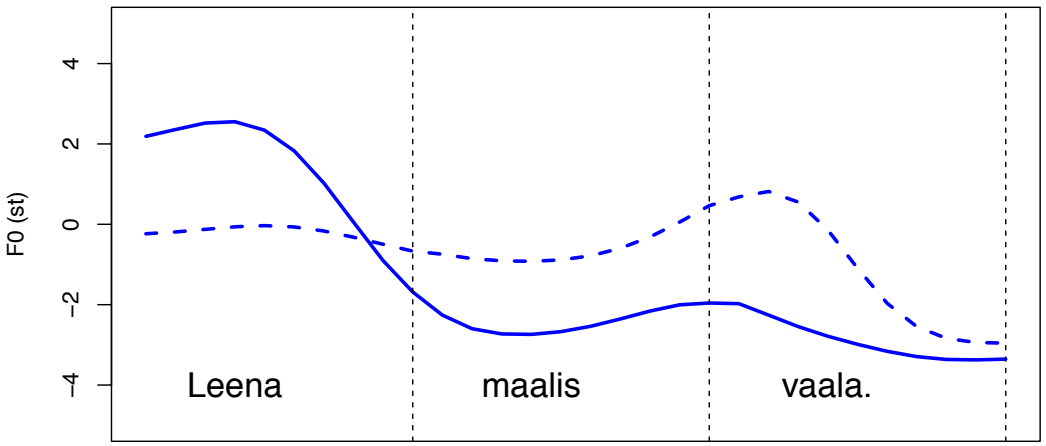

Figure 5.1. F0 contour of an utterance *Leena maalis vaala* ('Lena drew a whale') with focus either on *vaala* (dashed line) or on *Leena* (solid line) aggregated over 17 speakers as a function of normalized time (10 F0-values in equal time-intervals for each word in the phrase).

Another hypothesis that Terken and Hirschberg (1994: 140) put forward is that a grammatical function either supports pitch-accenting or interrupts the predicted deaccentuation. Their proposal is connected to the statistical relationship between grammatical function (subject noun phrase *vs*. object noun phrase) and information-structural implication. The subject noun phrase is more frequently given information and, therefore, usually deaccented, whereas the object noun phrase is more frequently new information and, therefore, usually pitch-accented.

Terken and Hirschberg (1994) did not find any evidence for their hypotheses. They suggest that it is due to strong interaction between sentence position and grammatical function in English. In English, the position in an intonational phrase varies together with the grammatical function: subject noun phrase is always in a prenuclear position of a phrase. Estonian, as learned in the previous chapters, is a language that enables relatively free order of sentence constituents (Lindström, 2004; 2006) while the grammatical functions, like subject or object, are assigned by different morphological forms (nominative case *vs.* partitive/accusative case). Thus, the sentences are grammatical in OVS as well as in SVO word order and can be interpreted

as neutral to focus as well. Therefore, the effects of position and grammatical function on deaccentuation could be tested independent of each other.

5.2. Experiment

In order to test the effects of sentence position and grammatical function, an experiment was designed in which native Estonian speakers were asked to utter sentences written on a screen in a natural and communicative way.

5.2.1. Materials

The sentences were transitive sentences consisting of a verb *kuulama* (‘to listen’) and two *sentence arguments*: an object (O) and a subject (S) noun phrase. One of the sentence arguments was treated as a *target word* that was a noun referring to one of the animate human entities: *beebi* ‘baby’, *diiva* ‘diva’, *joogi* ‘yogi’, *laama* ‘lama’, *leedi* ‘lady’, *liige* ‘member’, *muusa* ‘muse’, *piiga* ‘little girl’. The target word was placed into the sentence-initial or the sentence-final position; it was either a subject or an object in those different sentence positions. This resulted in two types of word orders in the materials: SVO and OVS.

All target words consisted of a long vowel that occurred in the first stressed open syllable and a short vowel in the second unstressed syllable. The nouns were given in nominative form, which is also a case for subject or semantic agent in Estonian. If the target word is the object or semantic recipient, then it is in the partitive/accusative case with a morphological marker *-t*. Thus the objects in the experimental material consisted of a stressed open long syllable and an unstressed closed long syllable and are therefore intrinsically longer than subjects (*beebi vs*. *beebit*). As for the quantity, both the subjects and objects were in long quantity (Q2, see section 1.1.4 for further details on Estonian three-way quantity system). The other sentence argument called as *non-target word* was a proper noun that also contained a long vowel in the first stressed syllable and a short vowel in the second unstressed syllable (*Jaana, Leena, Liina, Loona, Riina, Taavi, Tiina, Viive*).

Givenness was defined as repetition of a sentence constituent. The list of sentences was constructed in such a way that the sentence consisting of a word that was

the target of the measurements (target word) preceded different sentences consisting of this word (see example 5.1). The assumption was that with each repetition, the target word became more activated for the speaker.

(5.1)

a. [Diiva]$_{\text{TARGET}}$ kuulab [ööbikut]$_{\text{NON-TARGET}}$. (A diva listens to a nightingale.)

b. [Diiva]$_{\text{TARGET}}$ kuulab [õpetajat] $_{\text{NON-TARGET}}$. (The diva listens to a teacher.)

c. [Diiva]$_{\text{TARGET}}$ kuulab [presidenti] $_{\text{NON-TARGET}}$. (The diva listens to a president.)

d. [Diiva]$_{\text{TARGET}}$ kuulab [Taavit] $_{\text{NON-TARGET}}$. (The diva listens to Taavi.)

Observe in (5.1a) that the noun phrases *diiva* ('a diva') and *ööbikut* ('a nightingale') are not mentioned previously. Therefore, the target (*diiva*) and the non-target word (*ööbikut*) in the first sentence are both information-structurally NEW (they appear without an influence of previous context). However, in (5.1b, 5.1c and 5.1d) the target word *diiva* is already mentioned in the previous sentence and, therefore, it is activated in a discourse. Thus, the target word *diiva* is GIVEN and the non-target word *Taavit* in FOCUS in (5.1d). In addition to the main research question, it is hypothesized that the repetition of the target word will show gradual effects in F0 compression: the second repetition in (5.1c) is weaker than the first (5.1b) and the third repetition (5.1d) is weaker than the second (5.1c). The mentioning of a target word is called *occurrence*, including the first mentioning and the three repertitions.

Blocks of sentences similar to the example in (5.1) referred to as *target blocks* were constructed. Target-blocks were alternated with analogous blocks with filler sentences – *filler blocks*. Filler blocks also consisted of three-word-sentences but in passive voice. They consisted of an object and an adverb and were segmentally and rhythmically less controlled. In order to distract speakers from four-sentence blocks, the number of sentences in filler-blocks was varied between 3 and 4.

In total, the materials consisted of 64 target sentences 2 (context: given, new) * 2 (position: final, initial) * 2 (grammatical function: object, subject) * 8 different target words (items) that were presented to two groups of listeners, so that each participant saw the target sentence only in two conditions (for example, only as given initial object and as given initial subject noun phrase). Two lists of sentences were created. The list of sentences consisted of 16 target-blocks (instead of 32) and 25 additional filler-blocks.

5.2.2. Procedure

The participants were seated in front of a computer screen in a soundproof booth located at the University of Tartu in Estonia. The sentences were presented one by one after a mouse click by the participant. The participants were asked to first read the sentence and then utter it in a natural way as if they were speaking to a friend. They were advised to memorize and compare the current sentence with the preceding sentence while performing their task. They could proceed at their own pace. The sentences were presented as a slide presentation in Praat (Boersma & Weenink, 2014). In order to avoid order effects, each participant was presented with a different randomized list of the blocks.

5.2.3. Participants

The participants were 30 female and 14 male speakers (44 altogether) between 20 and 47 years of age (mean = 27.4 years) with normal hearing. The speakers originated from different dialectal areas all over Estonia: 18 from Northern Estonia, 14 from Tallinn; 15 from Southern Estonia, 11 from Tartu; 11 from Western Estonia and 2 from Saaremaa. One of them reported to have lived in Germany more than 3 years and to have a high proficiency of German; one of them reported to have very good knowledge of Russian and one of them a good knowledge of French. They were either students of the University of Tartu or professionals from Tartu and Tallinn. The participants contributed voluntarily.

5.2.4. Analysis

The expected amount of data was 1408 utterances: 44 (participants) × 4 (conditions for each participant) * 8 (items); the acquired amount of data was 764 utterances. For some participants, the reason for excluding quite a number of sentences from the analysis was a difficulty to utter written sentences in a natural way. Also, the utterances with list intonation were excluded from the analysis. Unfortunately, the experimental design supported the list intonation quite strongly (consider (5.1) with this respect in mind). Based on the previous results (chapters 2 and 3), the fourth sentence was expected to have a clear prosodic focus in the non-target position. Therefore, all the productions without a clear prosodic prominence on the word in focus were omitted. Additional

analysis is needed, but a great number of excluded utterances consisted of three downstepped pitch accents that sounded equal in prominence. In addition, because the F0 could not extracted reliably, a number of utterances was excluded due to a *creaky voice*. Creaky voice has reported to be quite widespread in Estonian spontaneous speech (Aare, 2014).

The words and sounds were force-aligned (using software provided by Arumäe, 2014); F0 contour was manually annotated, relying on perception and visual observation of the F0 track in Praat (Boersma & Weenink, 2014).

In the sentence arguments that were target or non-target words, the salient starting point and end point for the F0 excursion called elbows was determined by visual observation (as section 3.2.4). The first elbow (F01) was annotated in the vowel of the stressed syllable and the second elbow (F02) following the first one somewhere at the syllable boundary or in the second unstressed vowel. If the pitch was falling, the first elbow was high; if it was rising, the first elbow was low (see Figure 5.2 for illustration).

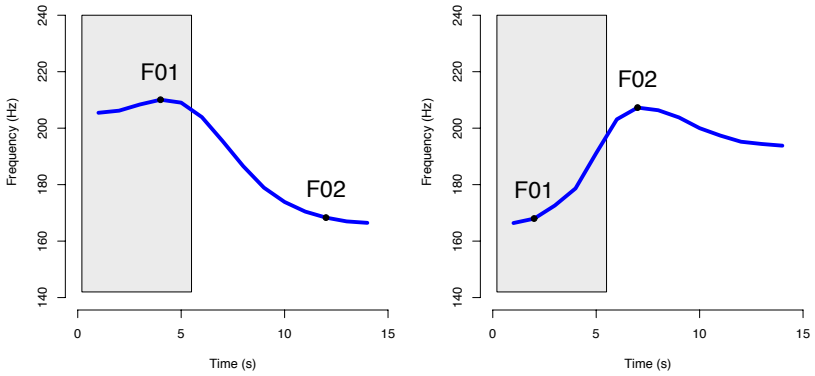

Figure 5.2. Manual annotation of the F0 curves in the disyllabic target and non-target words. The grey box represents the mean duration of the first stressed syllable. F01 represents the first elbow and F02 the second elbow.

Not all the instances were conspicuous falls or rises as the ones in Figure 5.2. If F0 was flat and no clear F0 maximum or elbow could be detected, theoretical elbow was annotated. The theoretical elbow was determined on the basis of the theory of Estonian quantity system (see chapter 1.1.4, or Lehiste, 1960; 1997; Lippus et al., 2013; Mihkla and Kalvik, 2011). According to the theory, in words in long quantity (Q2), the

peak is aligned with about three-quarters of the length of the first stressed vowel and the target of the changing F0 is reached at the beginning of the second syllable (Lehiste, 1960; Eek, 1983; Lippus et al., 2013). Therefore, in case of a flat F0 the first elbow was determined to be in the three-quarters of the first vowel and second elbow in the beginning of the second vowel.

F0 was converted into semitones using (5.2):

(5.2) $$\text{Semitone (st)} = 12 * \log(F0/F0_{spMean})$$

The speaker mean was calculated as the mean value of all the F0 samples generated by Praat pitch analysis (default time step, pitch floor 75 Hz) from all the utterances; see chapter 3 for further details.

High elbow was interpreted and analysed as *peak height*.

Slope was defined as F0 change divided by its duration (see the formula in 5.3).

(5.3) $$\text{Slope (st/s)} = (f0_{F01} - f0_{F02}) / (T_{F01} - T_{F02})$$

In this experiment, F0 slope estimates the degree of F0 compression. F0 slope greater than zero means the F0 excursion of large magnitude (the word is pitch-accented); F0 slope close to zero means the flat F0 excursion (the word is deaccented). Slope is negative, if measured for rise (see the right panel of Figure 5.2). Absolute value of the slope is going to be plotted and evaluated. The experiment assumes that any kind of F0 protrusion (fall and rise) elicits prominence and should not be considered as deaccentuation. The two different directions of F0 excrusion might have different pragmatic meanings, but this is not considered in this study.

Vowel duration in the nucleus of the first stressed syllable was measured as an additional parameter for the segmental suppression of given expressions shown to occur in the previous studies of other languages (Cooper et al., 1985; Eady & Cooper, 1986; Fowler, 1988; Bard et al., 1995; Fowler et al., 1997). Vowel duration was measured instead of syllable or word duration for two reasons. First, the segmental structure of the target words was CVVCV(C), thus the duration of the syllable happens to be mainly the duration of the long vowel. Second, the word duration was inappropriate, because the objects had the case marker *-t* at the end of the word, which makes them longer than the subjects that had no marker at the end. Word duration would result in an uninteresting effect of the presence of morphological case marking.

Peak difference between the peaks on the sentence-initial and sentence-final noun phrase was calculated. As in previous chapters, peak scans the shape of the topline. With regard to topline, it is remainded that the F0 peak is defined differently for the falls and rises: for the fall, it is high elbow followed by a fall (F01); and for the rise, it is high elbow (F02) preceded by a rise. A peak difference close to zero means that the difference in height between F0 peaks is small, and that both peaks are at the same F0 level in a phrase. Peak difference clearly greater than zero means that the first F0 peak is higher than the second F0 peak. Peak difference of a clearly negative value means that the first F0 peak is much lower than the second F0 peak.

The effects of givenness, sentence position and grammatical function were estimated in the generalized linear mixed models separately for each dependent variable peak height, F0 slope, vowel duration and peak difference (as a method available in the lme4 package (Bates et la., 2012) in software R ("R Development Core Team", 2014). Subjects and items were set as random factors. P-values were obtained by likelihood ratio tests of the full model with an interaction in question against the model without the interaction.

Next section first examines the effect of repetition on prosody (5.3.1), then the two types of pitch excursions (5.3.2) and finally the effect of givenness on intonation in relation to position and grammar (5.3.3).

5.3. Results

5.3.1. Prosodic effects of repetition

As explained in section 5.2.1, every participant uttered blocks of 4 sentences each. The following section investigates whether three repetitions of a target word had an effect on peak height (st), slope and vowel duration. Degree of activeness of the expression is hypothesized to increase as a function of repetition of the target word. For this, a gradual decrease of F0 excursion and vowel duration is expected.

Figure 5.3 demonstrates the effect of repetition on peak height.

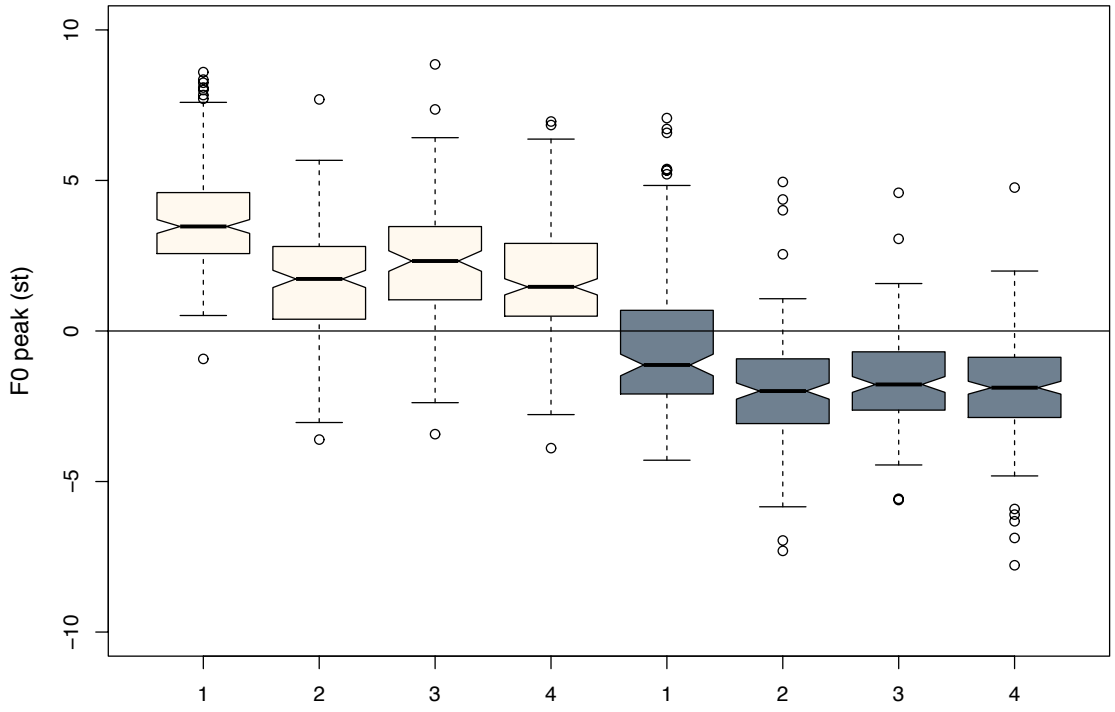

Figure 5.3. F0 peak (st) as a function of occurrences (1 – 4). The white boxplots show the target words placed sentence-initially and the grey boxplots the target words placed sentence-finally.

Generalized linear mixed model with the dependent variable *F0 peak*, with fixed effects position (initial, final) and occurrence (1,2,3,4) and with random effects subject and item showed a significant interaction between position and occurrence ($\chi 2[1] = 8$, $p < 0.05$). Post-hoc Tukey comparisons revealed that in both positions the first occurrence is significantly ($p < 0.001$) different from the first, second and third repertition (2., 3., 4. occurrence respectively), but there were no significant differences between the first and second or the second and the third repetition. Thus, it appears that the peak does not get lower after the second occurrence and the repetitions do not differ from each other.

Figure 5.4 shows F0 slope in relation to occurrence and position.

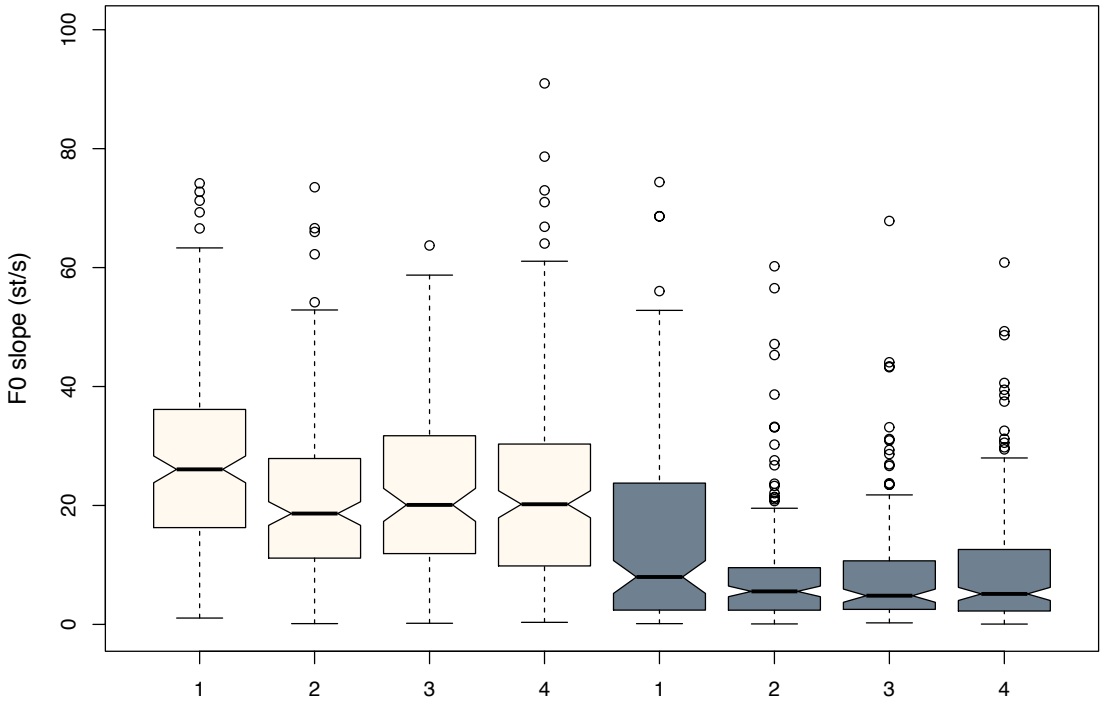

Figure 5.4. F0 slope (st/s) as a function of occurrences. The white boxplots show the target words placed sentence-initially and the grey boxplots the target words placed sentence-finally.

Observe in Figure 5.4 that the slopes in sentence-initial position are greater than the slopes in sentence-final position. In the sentence-initial as well as in the sentence-final position, the slope of the first mentioning is always greater than the slopes of repetitions. The slopes across the repetitions do not differ from each other.

Generalized linear mixed model with the dependent variable *F0 slope* and with fixed and random effects as above showed significant main effects of position ($\chi 2[1] = 328$, $p < 0$) and of occurrence ($\chi 2[3] = 75$, $p < 0$). Post-hoc Tukey comparisons showed that in both positions the first occurrence was significantly ($p < 0.001$) different from the second, third and fourth, but there were no significant differences between the second and the third or the third and the fourth occurrence. This means that the hypothesis did not gain any support and that the degree of F0 compression does not increase with the repetition of a lexical item.

Figure 5.5 shows vowel duration as a function of occurrence and position. Observe in Figure 5.5 that the vowel is shorter in sentence-initial than in sentence-final position. The vowel appears to be longer in the first occurrence than in the following repetitions for sentence-initial as well as for sentence-final position.

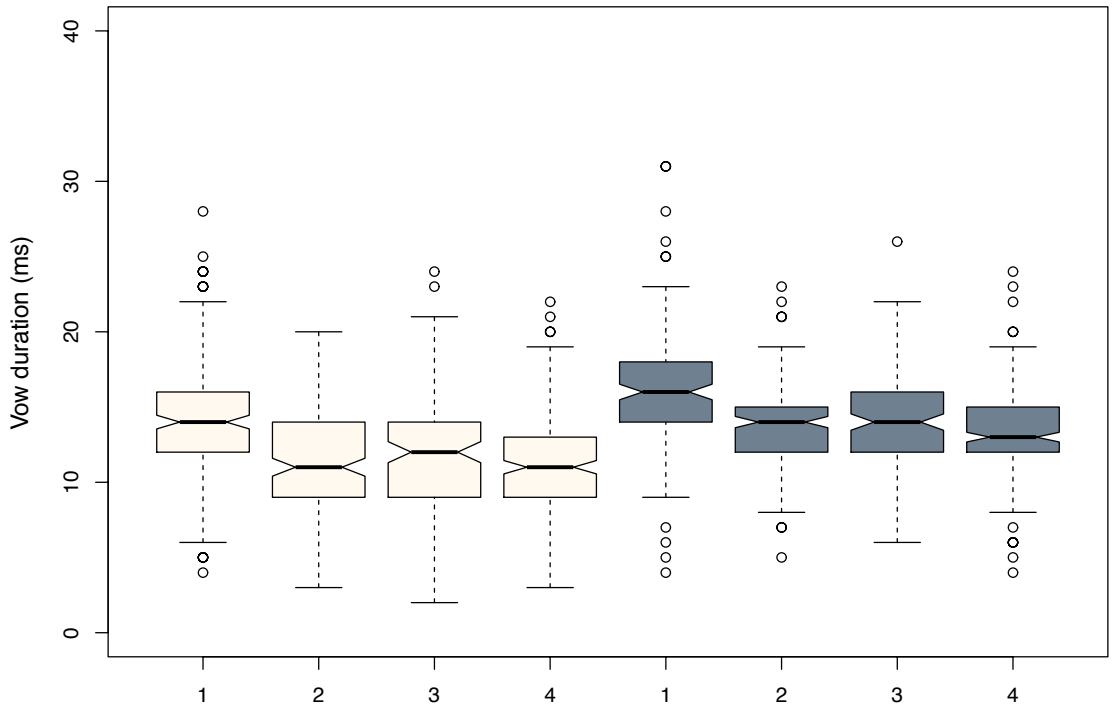

Figure 5.5. Vowel duration (s) as a function of repetition (1–4). The white boxplots show the target words placed sentence-initially and the grey boxplots the target words placed sentence-finally.

Generalized linear mixed models with the dependent variable *vowel duration* and with fixed and random effects as above showed significant main effects of position ($\chi 2[1] = 166.2$, $p < 0.001$) and of occurrence ($\chi 2[3] = 120.16$, $p < 0.001$). Post-hoc Tukey comparisons showed that in both positions the first occurrence is significantly longer in vowel duration than the first, second and third repertition ($p < 0.001$), but there is no significant differences between the repertitions. Thus, there was no difference in vowel duration across the repetitions.

To sum up, the analysis of repetitions showed that the repetition did not cause any increase in compression of F0 range or segment duration. Therefore, for investigation of the effect of givenness, it was quite arbitrarily decided to investigate the first and the fourth occurrence of the target word.

5.3.2. Analysis of falls and rises

Difference between F0 peak of the first elbow (F01) and the second elbow (F02) (see Figure 5.2 for reference) was calculated as a range of F0 excursion. A range greater than zero was defined as a fall; a range smaller than zero was determined as a rise. In total, there were 523 pitch falls and 241 pitch rises. Values very close to zero reflect a flat F0 contour. The following section examines phonetics of falls and rises.

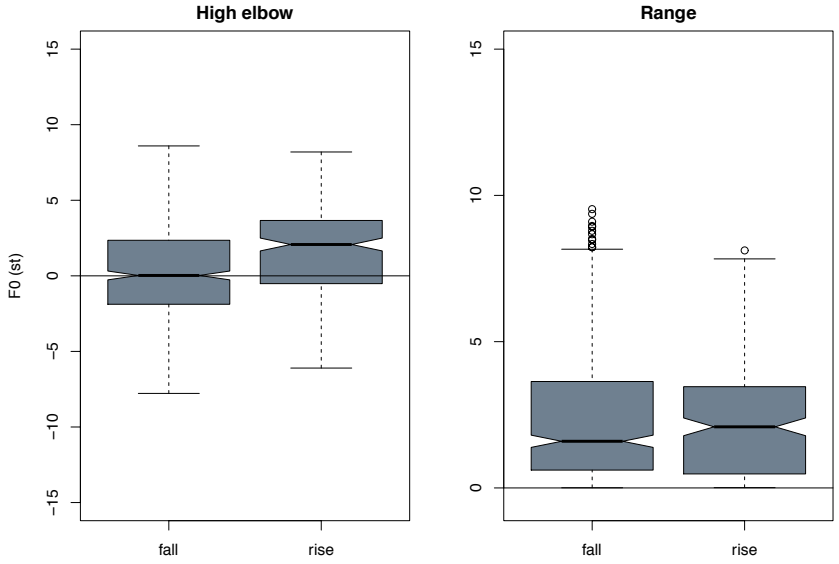

Figure 5.6. F0 peak and range of the F0 excursion (in semitones) in falls and rises.

A pitch maximum as well as a high elbow (see Figure 5.2) was interpreted as peak. On the left, Figure 5.6 shows that the peak of a fall is lower (it is closer to zero-line) than the peak of a rise. The right panel of Figure 5.6 shows that the range of F0 excursion does not differ between falling and rising excursions.

Figure 5.7 plots low and high elbows as a function of fall and rise.

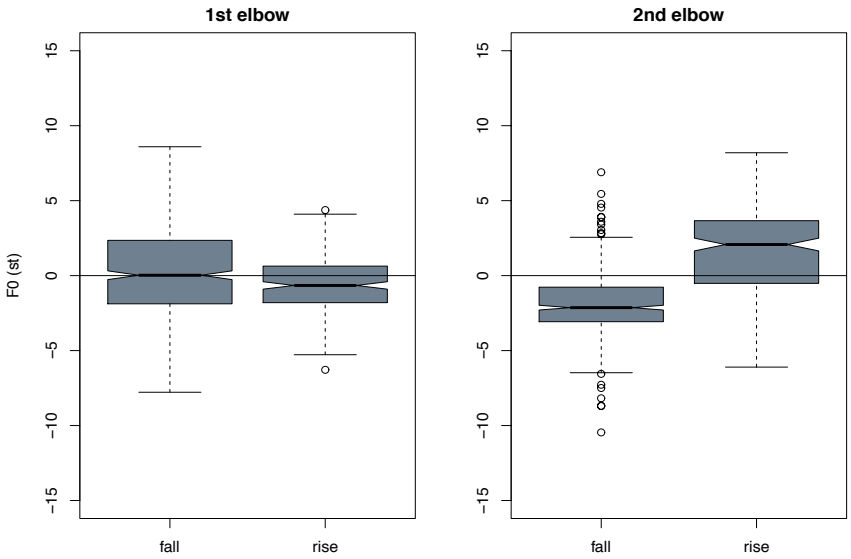

Figure 5.7. First (on the left) and second elbow (on the right) in falls and rises.

The left panel of Figure 5.7 demonstrates that a pitch excursion started at about the same level in falling and rising excursions. The boxplots of the second elbow in the right show that while the pitch dropped below the speaker mean (zero-line) in falling excursions it raised higher than the first elbow in the falling excursion. Figure 5.6 showed that the range of excursion was the same in falls and rises. Thus, the F0 dropped and raised within the same pitch range, but since the drop or the rise starts from about the same level, the peak in the rise is higher than in the fall, which is a very interesting result.

Figure 5.8 plots the time of the first and second elbow proportionally to vowel duration that was calculated as shown in (5.4):

(5.4) Proportional time= $(t_{f0}-t_{vOn})/t_{vOff}-t_{vOn}$

Figure 5.8 demonstrates that the timing of the peak in the rise and the fall is different: in the falling excursion, the peak is followed by pitch fall and is therefore aligned earlier with the vowel onset, whereas in the rising excursion, the peak is preceded by pitch rise and therefore is aligned later with the onset.

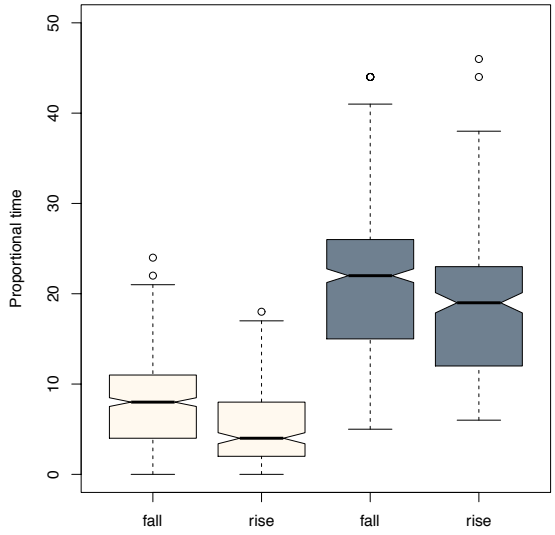

Figure 5.8. Time proportionally to the vowel duration of the first elbow (white) and the second elbow (grey) in falls and rises.

Observe in Figure 5.8 that, interestingly, both elbows occurred earlier for the rise than for the fall. Thus, the start of a rising excursion was closer to the vowel onset than the start of a fall.

Figure 5.9 plots the duration of the two excursion types. It can be observed that falls and rises did not differ in duration.

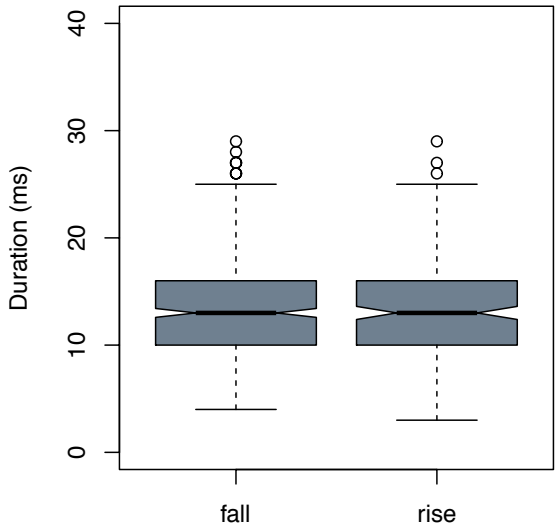

Figure 5.9. Durations (ms) of the falls and rises.

To sum up, it was observed for two types of excursions that they differed in alignment in relation to the stressed vowel. The rise started and ended slightly earlier than the fall. Observe in Figure 5.6 that the first elbow of both excursions was located at about 10% in the vowel. Therefore, the rise could be represented phonologically as low-rising pitch accent (L*+H) and the fall as high falling pitch accent (H*+L). Interestingly, the peak (H) in rise was higher than in fall. More importantly, the rising and falling excursions did not differ in range and duration. Therefore, the F0 slope would be the same size for both excursion types and this justifies pooling the falls and rises together for the investigation of pragmatic factors in the following section. The peak is the high elbow and the slope the absolute value of positive (fall) and negative slopes (rise).

5.3.3. Deaccentuation: pre-focal vs. post-focal position

This section tests whether givenness, defined as repetition of a lexical item in a particular discourse, compresses F0 excursion? Second question is whether this compression is affected by a grammatical function or by a position of a referring expression in a sentence. The degree of F0 compression is quantified in peak height and F0 slope. In addition to F0 compression, also duration of the stressed vowel is examined. For capturing the tonal characteristics of the whole sentence, the peak difference between the peaks was calculated. Table 5.1 presents the number of observations across the experimental conditions.

Table 5.1. Number of observations.

		Information structure	
Position	Grammar	Given	New
Initial	Object	92	96
	Subject	103	106
Final	Object	76	114
	Subject	73	104

Figure 5.10 shows the time-normalized F0 contours averaged over 22 speakers and 8 conditions. The plots at the top show SVO and at the bottom OVS word order. In the right-hand plots the target word was located at the beginning of the sentence; in the

left-hand plots at the end of the sentence. The effect of givenness is observable in F0 contours in black.

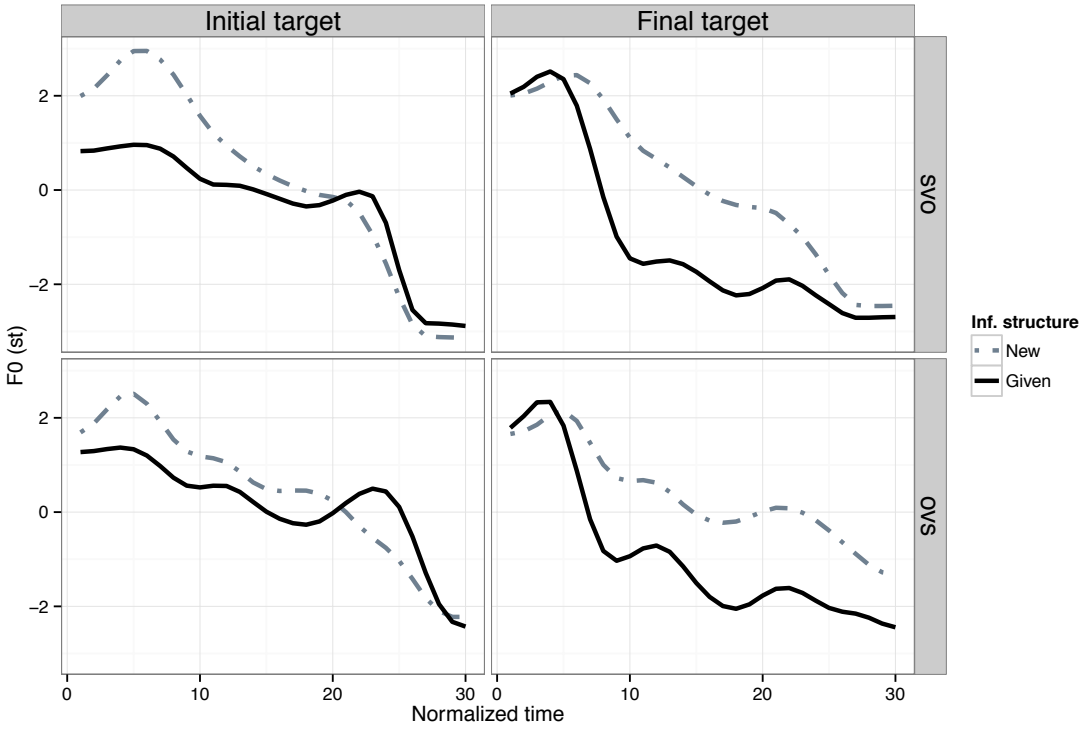

Figure 5.10. F0 contours (in semitones) aggregated over 22 speakers as a function of normalized time (10 F0-values in equal time-intervals for each word in a phrase). The rows show the contours from two word orders (SVO *vs.* OVS), the columns the location of the target word (initial *vs.* final) and the line types the information structure (new *vs.* given).

In Figure 5.10, it can first be seen that the contours do not vary with word order, which indicates that there was probably no effect of grammar. Second, the target-word in the initial position is carrying a pitch peak in the new as well as in the given condition, but the peak in the given condition is scaled considerably lower than in the new condition. Third, the pitch excursion on the target-word at the end of the phrase is very small or almost non-existent for the given condition. This indicates a clear effect of givenness. Fourth, it appears that givenness considerably affects the topline slope.

Figure 5.11 plots the peak height (st) in relation to information structure (new *vs.* given), grammar (subject *vs.* object) and position (initial *vs.* final).

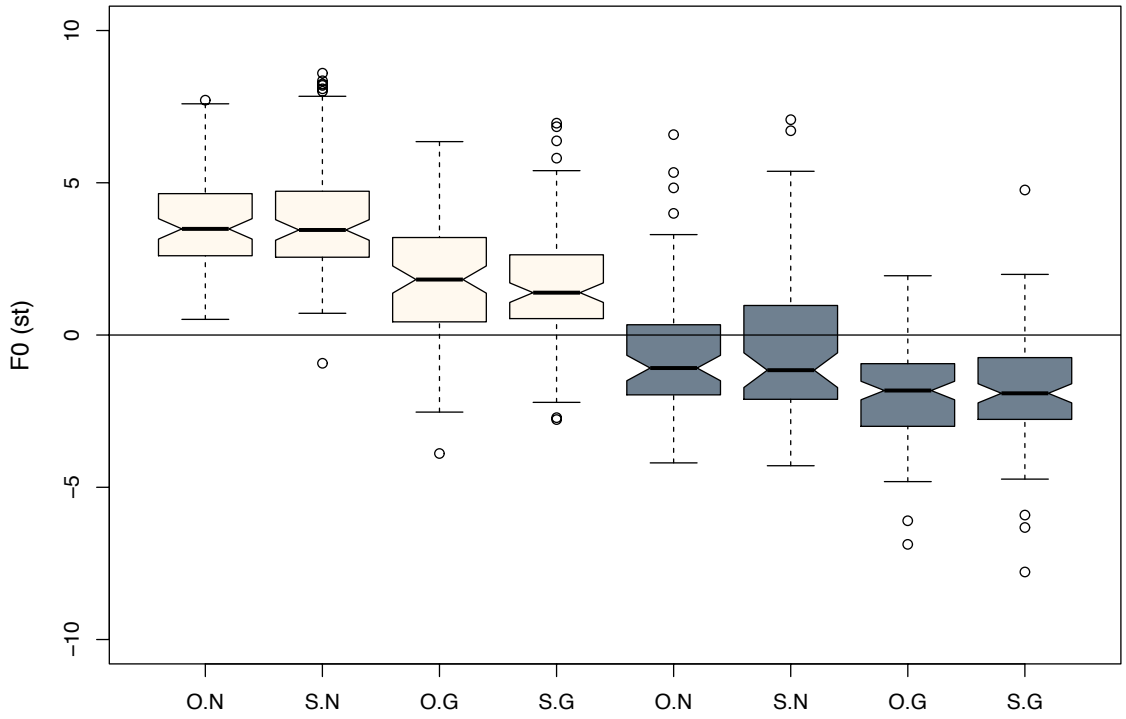

Figure 5.11. F0 peak measured in the stressed syllable of the target word in sentence-initial (white) and in sentence-final position (grey) in relation to information structure (N refers to New, G refers to Given) and grammatical function (O refers to object, S to subject noun phrase).

Generalized linear mixed models with the dependent variable peak height, with fixed factors information structure, grammar and position and with random factors subject and item[23] showed no significant interactions; there were significant main effects of position ($\chi^2[1] = 490$, $p < 0.001$) and information structure ($\chi^2[1] = 116$, $p < 0.001$).

In Figure 5.10, the peak in the sentence-initial position is considerably higher than the peak in the sentence-final position. Within position, the peak is lower for given than for new. There is no effect of grammar. Interestingly, though, the peak in the sentence-final new target word is also considerably low – below the zero-line.

Figure 5.11 plots absolute of F0 slope in connection to information structure, grammar and position. To recall, slope is a measure of F0 compression or F0 expansion in the target words: the greater the slope, the lesser the F0 compression; the closer the slope is to zero, the greater the F0 compression.

[23] lmer(peak_st~grammar * is * pos + (1|subj) + (1|target), data = df)

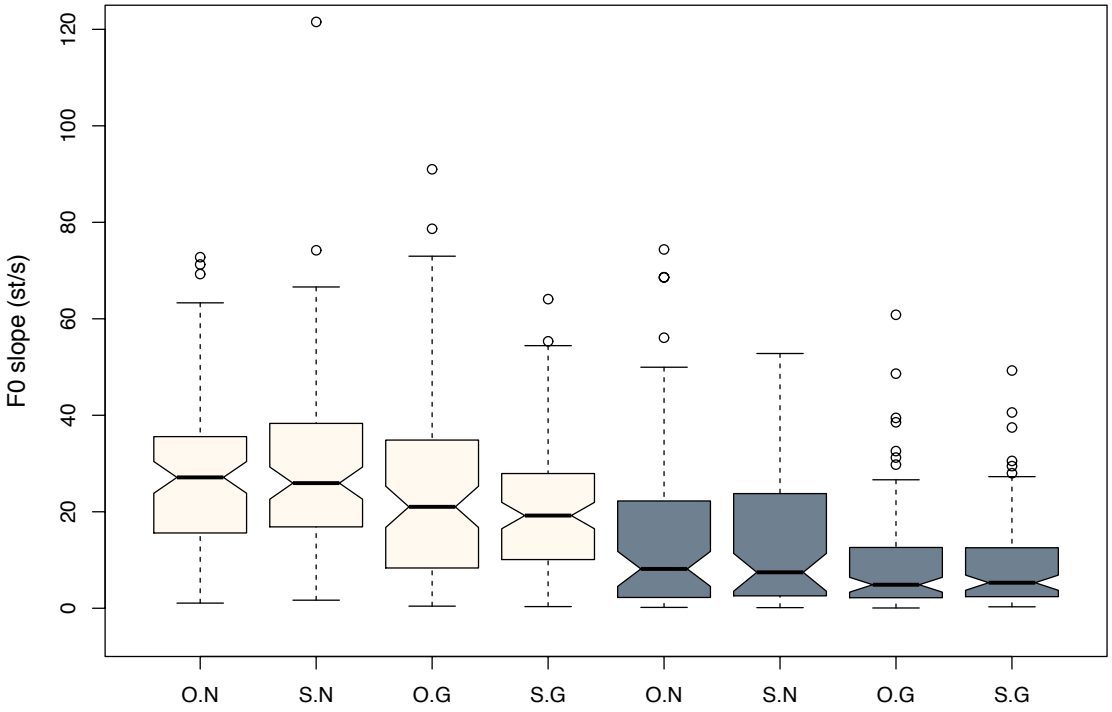

Figure 5.12. F0 slope measured in the stressed syllable nucleus of the target word in sentence-initial (white) and in sentence-final position (grey) in relation to information structure (N refers to New, G refers to Given) and grammatical function (O refers to object, S to subject noun phrase).

Generalized linear mixed models with the dependent variable F0 slope and with fixed and random effects as above[24] showed no significant interactions; there were significant main effects of position ($\chi^2[1] = 169$, $p < 0.001$) and information structure ($\chi^2[1] = 57$, $p < 0.001$). For the main effect of position, observe in Figure 5.12 that the slopes for sentence-initial are greater than for phrase-final. For given, the slope is close to zero in the phrase-final, but somewhat greater than zero in the phrase-initial position. This result gives confirmation to the hypothesis that deaccentuation depends on the position of the given word in a phrase. There is no effect of grammar. Interestingly again, the F0 slope of the sentence-final new target word is quite small (less than 20 st/s), whereas the sentence-initial new target word is considerably greater.

For duration analysis, the vowel duration was more appropriate than word duration, because the word duration varied due to absence or presence of the object-

[24] lmer(slope ~ grammar * is * pos + (1|subj) + (1|target), data = df)

marking *-t* at the end of the word, which is not interesting in regard to givenness (see section 5.2.1). Figure 5.13 plots the duration of the vowel in the stressed syllable.

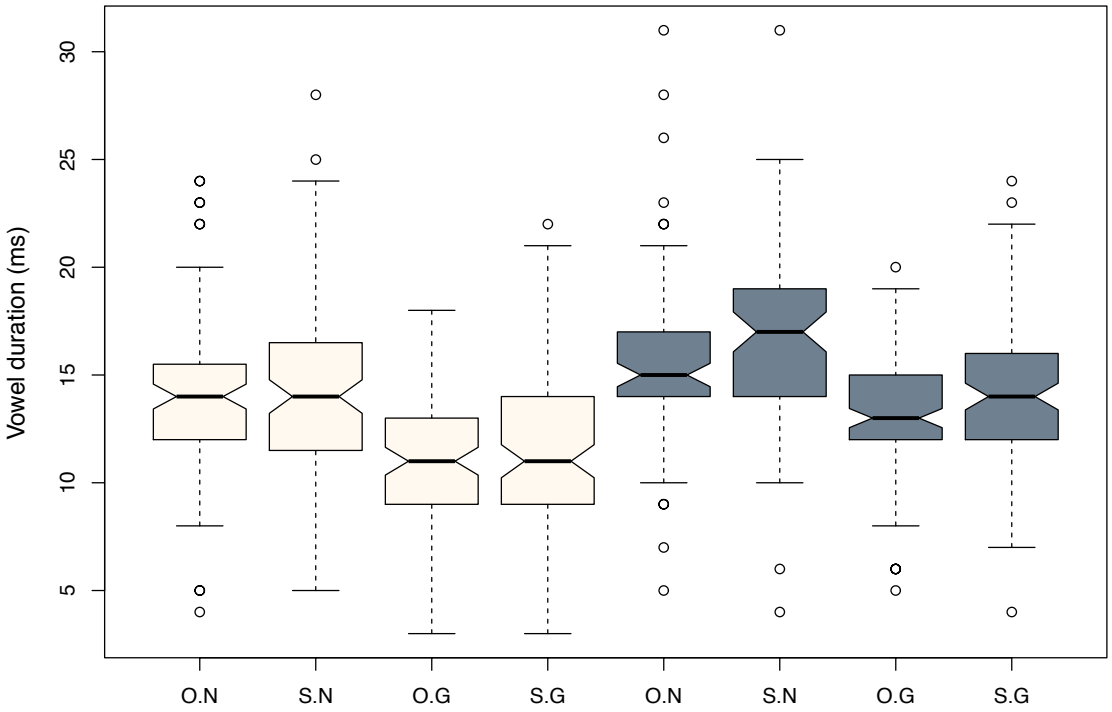

Figure 5.13. Duration of the vowel in the stressed syllable nucleus of the target word in sentence-initial (white) and in sentence-final position (grey) as a function of information structure (N refers to New, G refers to Given) and grammatical function (O refers to object, S to subject noun phrase).

Generalized linear mixed models with the dependent variable *vowel duration* and with fixed and random effects as above[25] showed no significant interactions. There were significant main effects of position ($\chi^2[1] = 73,\ p < 0.0001$), of grammar ($\chi^2[1] = 12.3,\ p < 0.0001$) and of information structure ($\chi^2[1] = 82.6,\ p < 0.0001$). In Figure 5.13, it can be observed that the vowel is shorter in the phrase-initial than in the phrase-final position. The same effect of information structure can be seen in both positions: given is shorter than new. The effect of grammar is difficult to trace in Figure 5.13: the vowel duration for subject in the sentence-final position appears to be slightly longer than for object.

[25] lmer(vDur~grammar * is * pos + (1|subj) + (1|target), data = df)

As next, peak difference is examined. Peak difference detects how many pitch accents in a phrase there were, and estimates pitch height of the given in relation to new. Table 5.2 presents the size of the data investigated.

Table 5.2. Number of observations in the analysis of declination.

		New	Given
Initial	Object	51	72
	Subject	63	76
Final	Object	39	117
	Subject	41	106

Figure 5.14 plots the peak difference measured in utterances with varying word order and information structure.

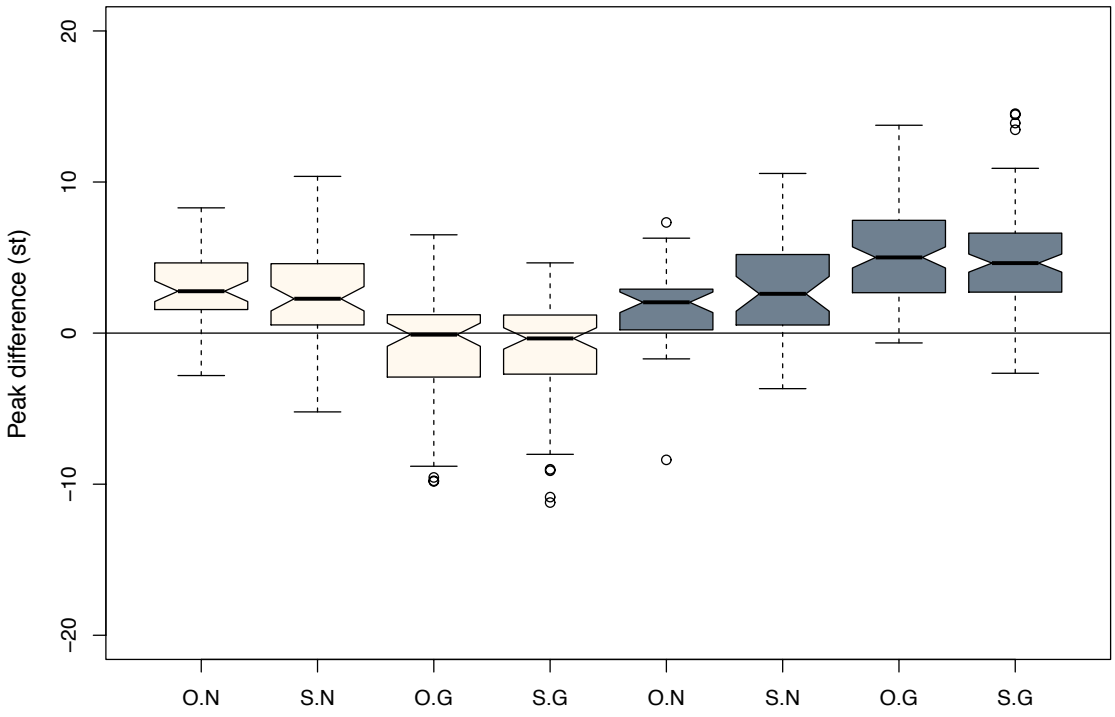

Figure 5.14. Difference between the F0 peaks in the sentences in which the target word occurred either sentence-initially (white) or sentence-finally (grey). The target word was either new (N) or given (G) in relation to the preceding sentence, and either subject (S) or object (O) noun phrase. Figure is to be interpreted as follows: the grey boxplot of O in the new condition means that the target word is a object noun phrase of the SVO sentence; the grey boxplot of S means that the target word is a subject noun phrase of the OVS sentence.

Generalized linear mixed models with the dependent variable *peak difference*, with fixed factors position (final *vs.* initial), grammar (O *vs.* S) and information structure (given *vs.* new) and with random factors subject and item[26] showed a significant interaction between the factors ($\chi 2[1] = 4.5,\ p < 0.05$). Information structure affected the peak difference in the sentence-initial as well as the sentence-final object ($p < 0.001$), and in the sentence-initial as well as the sentence-final subject ($p < 0.001$). The position affected the peak difference in the given object as well as in the given subject noun phrase ($p < 0.001$), but not in the object or the subject noun phrase carrying new information.

The first important result was that there is no difference between the sentence-initial subject and object for new information. This result suggests, in contrast to the result in chapter 3, that there is no difference between SVO and OVS word order. This contradicting result might be attributed to the design of materials that was a list of sentences.

Peak difference varied between 4 and 6 st when the contextually given target word was sentence-final. Such a great difference shows that the pitch on the sentence-final given word is very low in relation to the pitch on the sentence-initial new word. Peak difference is close to zero in cases where the given word is sentence-initial. This indicates that F0 peaks were at the same heights in the sentence, and that there was most probably a prenuclear pitch accent on the sentence-initial given word.

5.4. Discussion

A speech production experiment investigated for Estonian whether words containing given information are deaccented at the same degree at beginning and end of the phrase in the position of subject or object noun phrase. Based on the studies of English, it is often assumed that given information in a discourse is deaccented (see about deaccentuation in English in Halliday, 1967a; Brown, 1983; Ladd, 2008). Cruttenden (2006) in his study shows that the strategy to deaccent given information is not always used in the languages of the world. For example, the expressions carrying given

[26] lmer(peakDif_st ~ grammar * is * pos + (1|subj) + (1|target), data = declData)

information are frequently accented in Greek and Italian, whereas in Spanish, Russian and Swedish they are accented most of the time. Gårding (1981) proposes for Swedish that expressions carrying given information are accented due to lexical pitch accent. This line of argumentation, however, cannot account for Romance languages that also appear to resist deaccenting (Swerts et al., 2002; Cruttenden, 2006; Ladd, 2008).

The Estonian three-way quantity system relies to a great degree on tonal cues (Lehiste, 1960; 1997; Lippus et al., 2009; 2013), which makes Estonian more similar to Swedish. The results of a corpus study in Lippus et al., (2013), however, indicate that the tonal cues of the quantity are not realized in the case of deaccentuation. Therefore, this study did not concentrate on the interaction between quantity and givenness. Only words in long quantity (Q2) were included in the experimental materials. Compatibly with the corpus study, carried out in Lippus et al., (2013), the results showed that it was not difficult to deaccent Q2 words. However, the effect of givenness on the tonal cues associated with quantity still deserves a closer examination in future.

The second question dealt with the effects of sentence position and grammatical function on the degree of deaccentuation. Terken and Hirschberg (1994) have suggested that givenness is not sufficient to condition deaccentuation of the expression. They (1994: 138) hypothesize that due to the rhythmical reasons, the givenness preceding the focus is signalled in smaller pitch range and not with the absence of pitch accent. They do not find any support for their hypothesis. In their study, speakers chose either "to accent or deaccent an expression" (Terken and Hirschberg, 1994: 138). They attributed their negative result to too few data. The hypothesis was re-tested with a prediction that in pre-focal position the pre-focal pitch accents are free to occur, but in post-focal position, givenness causes deaccentuation. The results confirmed the prediction: the slope (see Figure 5.12) for given expressions was considerably greater in sentence-initial position than in sentence-final position. However, the proportion of utterances (about 50%) excluded from the analysis might indicate that it was difficult for the speakers to produce clear deaccentuation. This in turn might mean that deaccentuation of the kind that occurs in Germanic languages is not very typical in a language like Estonian.

Predictably, when the target word was in sentence-final position, its pitch peak was typically lower than when it was in the initial position. Interestingly, the peak was low also in the sentence-final new condition. This effect is very likely a reflex of the F0

declination. F0 declination could also explain the observation that nuclear pitch accents are not acoustically as prominent as prenuclear position (Ladd, 2008: 259). However, the listeners are known to normalize the F0 declination in spoken languages in a way that a pitch accent in the sentence-final position is perceived as high as the pitch accent in initial or medial position even if the former is acoustically lower in pitch (Pierrehumbert, 1979).

Interestingly, slope of new word in the sentence-final position was also considerably smaller than in the sentence-initial position. This result therefore supports the observation about nuclear pitch accents – they are not necessarily produced as acoustically prominent (Ladd, 2008: 259). The question for the potential perception study would be, therefore, how does pre-pausal, in other words, utterance-final position, interact with the perception of prominence. The results of this study might indicate that a small pitch slope is perceived more prominent in sentence-final position than in medial or initial position.

The position of the word in a phrase influenced the declination slopes for new and given information significantly. First, when the given word was phrase-final, the declination slope was the steepest probably due to drastic post-focus F0 drop. Second, when the given word was phrase-initial, then the declination slope was close to zero, which indicates that the phrase-final nuclear pitch accent was preceded by the prenuclear pitch accent. Interestingly, the declination slope was about 1.7–3.0 semitones in both of the word orders for the condition of new information. This declination slope is greater than occurred in the experiment presented in chapters 3 of 4. This might be the effect of context. In chapters 3 and 4, the utterance was a response to a question. In the experiment here, the utterance was an utterance within a list of sentences with words that were repeated in regular intervals.

Terken and Hirschberg (1994: 128) propose that the grammatical function interacts with sentence intonation also. They support their proposal by theoretical consideration that there is a correlation between grammatical function (subject *vs*. object) and information-structural implication, e.g. the subject noun phrase contains given information more frequently, whereas the object noun phrase tends to contain new information. They propose that this information-structural implication related to grammatical function might interfere deaccentuation stipulated by the context. This proposal was decided to re-examine for Estonian, which is a free word order language,

and enables to examine the effects of sentence position and grammatical function independent of each other. The results showed that grammatical function (subject noun phrase *vs.* object noun phrase) did not influence the speakers' choice to deaccent.

This result, however, is interesting in the light of the results reported in chapter 3. The result on focus prosody showed that nuclear pitch accent occurred phrase-initially in OVS word order in the broad focus context (as an answer to the question *What happened?*). The conclusion in chapter 3 was that OVS word order is possible only with an accent on the object noun phrase. In the experiment presented in this chapter, one of the experimental conditions involved embedding a sentence with OVS word order into a context with unrelated information (it was the first utterance in a block of sentences (see the example in 5.2), in which it was always preceded by a filler sentence). In this context, all the constituents of the sentence were new to the speaker. It was not explicitly mentioned in section 5.1 but similar effect as in chapter 3 was expected again: phrasal prominence on the sentence-initial object noun phrase. Contrary to this expectation, peak difference did not show any difference between the two word orders. Both OVS and SVO word order had nuclear pitch peak at the end of the phrase.

An explanation would be that the intonation occurred might be the intonation of theme-rheme structure as define in Halliday (1967): a structural principle of a sentence that does not have any information-structural value (or any hint on the focus of a sentence). If assumed that the downstep does not have any information-structural meanings, then this would also explain the great number of utterances with downstepped pitch accents, which were excluded from the analysis.

The contradicting results, thus, could be attributed to the pragmatic difference between 'broad focus context' and 'new presentational context'. A question, even a general one such as *What happened?,* might encourage a speaker to give an utterance some kind of focus interpretation. Uttering new presentational information (theme-rheme structure), for example, the first introductory sentence of a story, does not appear to have this kind of encouraging effect. The study in chapter 3 indicates that OVS word order in Estonian implies the focus interpretation on the object noun phrase. However, the results of this study indicate that the two word orders (SVO, OVS) might also be interpreted neutral to the focus like proposed in Välimaa-Blum (1984; 1993) and in some contexts might be used just for the theme-rheme structure as identified by Erelt et al. (1993).

In the experimental design, givenness was defined as a repetition of a referring expression. The expression investigated was repeated 4 times in a block of 4 sentences (following the design from Terken and Hirschberg, 1994). As a matter of interest, section 5.3.1 explored whether the increasing number of repetitions caused greater F0 compression and shorter segment duration. There was no significant effect of repetition. This study did not investigate this, but it could be that full nouns or proper names in the experimental material might be less prone to tonal and segmental suppression. The frequency is known to mainly affect pronouns and short function words.

5.5. Conclusion

The speech production experiment investigated deaccentuation of given information in Estonian. Deaccentuation was established by a degree of F0 compression that was quantified by the F0 peak and slope.

The results showed that givenness in Estonian causes different degrees of F0-compression depending on the location of the sentence constituent in reference to the location of the sentence focus: the F0 excursion in the sentence constituent following the focus is severely compressed (deaccentuation), whereas the F0 excursion in the constituent preceding the focus is just scaled lower from the corresponding word carrying new information. Vowels turned out to be shorter in given condition than in new condition, but always longer in sentence-final position, probably due to phrase-final lengthening. There was no effect of grammar in deaccenting on a sentence constituent.

In general, the study demonstrates that in addition to the information-structural factors, the structural principles of intonational phrase (prenuclear *vs.* postnuclear) also influence the tonal means of givenness. Together with the studies reported in chapters 2, 3 and 4, the study demonstrates that Estonian, belonging to a group of free word order languages, uses tonal means for transmission of the information structure.

6. Summary

Linguistic transmission of information among other linguistic and pragmatic principles and processing factors is influenced by the *information structure*. The communication of new or relevant information requires presence of information that is *established* or *shared* between the interlocutors (Vallduví, 1993; Lambrecht, 1994). In many cases, an utterance makes reference to the established information while stating something new or informative. The established (old) information and new information constitute the information structure of an utterance. The choice of a linguistic structure reflects information status that different referents can have in connection to a particular discourse (Chafe, 1976; 1987; Halliday, 1967a; Prince, 1981; 1992; Lambrecht, 1994; Gundel, 1999).

The linguistic categories of information structure that were defined and investigated in the study are focus (Halliday, 1967a; Lambrecht, 1994; Gundel, 1999), givenness (Chafe, 1974; 1976; 1987; Baumann, 2006), activeness (Chafe, 1976; 1987; Lambrecht, 1994) and identifiability (Lambrecht, 1994). In the framework of this study, *focus* was defined as a linguistic category for information that the speaker wants to present as new and unpredictable (Halliday, 1967a; Lambrecht, 1994). *Activeness* and *identifiability* capture the pragmatic properties of the components in the minds of speech participants (Chafe, 1976; Lambrecht, 1994). Activeness involves the status of a referent in the speech participants' consciousness. A referent can become activated through mentioning, but textual or situational inference also has an ability to activate referents. Identifiability involves speech participants' knowledge about the referents. A listener might know/be familiar with a particular referent or might not be. Lambrecht (1994) defines activeness and identifiability as binary linguistic categories. In Chafe's (1976; 1987) approach, however, givenness/activeness is a continuum. Experimental research has adopted the concept of *givenness* (Baumann, 2006; Baumann and Grice, 2006; Baumann and Riester, 2012) and as such the term was applied in the experimental part of the study.

The formal means that are most frequently accounted for focus in literature are accent and word order (Lambrecht, 1994; Gundel, 1999). The formal means for activeness/givenness are lexical encoding (full noun *vs.* pronoun) and accent (Chafe, 1976; 1987; Lambrecht, 1994; Gundel, 1999) or different types of pitch accents

(Baumann, 2006; Baumann & Grice, 2006). The best-known grammatical means for identifiability is definiteness, but also word order or morphological markers can encode identifiability (Chafe, 1976; Lambrecht, 1994; Féry & Krifka, 2008).

The main grammatical means that the current study was interested in were accent and word order. In introduction it was attested that intonation languages (e.g. English, Dutch and German) appear to use pitch accent and accent shift for signaling focus, and deaccentuation for marking activated/given information (Halliday, 1967b; Ladd, 2008; Brown, 1983; Cruttenden, 2006; Krahmer & Swerts, 2001; Breen et al., 2010; Kohler, 1991; Féry & Kügler, 2008; Baumann et al., 2006; Baumann & Grice, 2006). In some non-configurational languages focus is signalled by word order (É. Kiss, 1995; Rizzi, 1997; Keller & Alexopoulou, 2001; Skopeteas et al., 2009). The two means of information structure have been sometimes treated as mutually exclusive within a language. For example, Lambrecht (1994: 240) suggests that word order is used for pragmatic functions similar to the functions of sentence accent. One of the aims in the introduction was to show that word order might be exploited for the information-structural categories slightly different from the categories that employ pitch accent. For example, word order was attested to signal identifiability/definiteness (on the basis of data in Välimaa-Blum, 1988; Titov, 2012; Lindström, 2002) or activeness (on the basis of data in Ward & Birner, 2004; Weskott et al., 2011). This discussion demonstrated that the grammatical means such as pitch accent and word order cover various pragmatic functions in a discourse and a language might have both means available for an expression of information structure.

If a language employs pitch accent as well as word order for marking information structure, the question arises, how both means interact with each other. Hungarian and Finnish provided two theoretical models for the interaction between word order and accentuation. In Hungarian, pitch accent highlights the focus of a sentence together with word order (Siptár & Terköczy, 2000; Varga, 2002). The constituent in focus occurs before the verb and the preverbal position attracts the sentence accent. In Finnish, word order variation for focus marking is optional (Välimaa-Blum, 1988; 1993; Vainio & Järvikivi, 2006; 2007). Välimaa-Blum (1993) in her model suggests that the pragmatic meaning of word order can be *overridden* by additional linguistic context (e.g. by adding linguistic material to a sentence or by pitch-accenting). Vainio and Järvikivi (2007) have found some evidence for this suggestion.

Also, studies of unrelated languages such as Greek, Italian, Spanish, and Georgian have reached similar conclusions (Keller & Alexopoulou, 2001; Face & D'Imperio, 2005; Skopeteas et al., 2009).

If the hypothesis of overriding is valid, the next question arises, what motivates the overriding? In this connection a closer look to the word order variation is necessary. For some of the free word order languages (Finnish, Russian, Estonian) it is known that, in addition to focus or topic, they also encode identifiability of a referent by the sentence position (Välimaa-Blum, 1988; Titov, 2012; Erelt et al., 1993; Lindström, 2002). For example, in Finnish (Välimaa-Blum, 1988) and in Estonian (Erelt et al., 1993) identifiable referents can optionally be located in the beginning of the sentence and the unidentifiable referents at the end of the sentence.

Another information-structural category has shown to influence word order in well-known intonation languages such as English and German. Ward and Birner (2004) account for the preposing that causes word order permutation of OSV (object-subject-verb) in English and OVS (object-verb-subject) in German. Preposing is licenced by a context of *partially ordered relation – poset* (Ward & Birner, 2004; Weskott et al., 2011). In other words, a poset relation activates a referent in a discourse. It can also be called as 'activated or evoked text-internally' (Chafe, 1987; Lambrecht, 1994). Thus, information-structural category activeness might cause object-initial word orders in English and German.

As discussed in the first chapter, focus is a linguistic category for information that the speaker arbitrarily presents as new (Halliday, 1967a; Lambrecht, 1994; Gundel, 1999). From this arbitrariness of focus it follows that the referents in focus are not necessary inactive or unidentifiable. In contrast, the referents in focus can occasionally be already activated in a discourse or identifiable to the hearer (Lambrecht, 1994; Gundel, 1999). In this case, they bear unpredicted sentence accent. The proposal for Estonian is that a similar kind of overlap between information-structural categories is at work, if intonation appears to override the pragmatic implication provided by word order.

Estonian is a free word order language that belongs to the Finnic subgroup of Finno-Ugric languages (Abondolo, 1998; Vilkuna, 1998; Viitso, 2003) and this makes it closely related to Finnish and less closely to Hungarian. Estonian encodes various aspects of information structure (focus, definiteness) within the word order of a

sentence (Tael, 1988; Erelt et al, 1993; Lindström, 2002; 2004). Erelt et al. (1993) provided an impressionistic observation that in addition to word order, prosodic prominence is also used for focus marking.

The idea in Lambrecht (1994) that the three grammatical components are in competition with each other motivated the designs of the experiments. In a way, information structure, word order and intonation were set into a competition in all four studies, in order to answer two questions. The first question asked whether information structure has tonal correlates in Estonian. Second goal was to investigate whether information structure or word order interacts stronger with sentence intonation. The following five sections briefly summarize the results of the experiments carried out and explore them in the context of theoretical considerations presented above.

6.1. Perception of accent shift in Estonian

Intonation languages show a strong correspondence between focus and prosodic prominence. With the change of the position of nuclear pitch accent focus of an utterance also changes (Gussenhoven, 1983b; Cooper et al., 1985; Eady & Cooper, 1986; Swerts et al., 2002; Breen et al., 2010). Impressionistically, this has been claimed also for Estonian (Erelt et al., 1993). In addition, studies suggest that the sentence-final position is a focus position for neutral new information (as opposed to the contrast or correction) in Estonian (Erelt et al., 1993; Lindström, 2006), which implies that the referring expression occurring at the end of the sentence is automatically interpreted as the sentence focus. Vainio and Järvikivi (2006) have shown for Finnish in a slightly different experimental design that the sentence-final position is a strong focus cue for some types of noun phrases (location adverb). The aim of the study was to establish whether pitch prominence corresponds to focus in perception of Estonian L1-speakers. In the experimental design, two focus cues were contradicted with each other: sentence-final position and the location of the nuclear pitch accent.

The results showed that the sentence-final position did not have any effect on the perception of focus. The position of nuclear pitch accent was interpreted as focus of an utterance. An accent shift was interpreted as a change in focus, similarly to intonation languages (e.g. English; Gussenhoven, 1983; Swerts et al., 2002). This result gave rise to two kinds of considerations. The first consideration is connected to a study carried

out by Vainio and Järvikivi (2006). Their stimuli consisted of a verb and two adverbs. Thus, in the focus perception, the position of an adverb against the position of another adverb was tested. The stimuli consisted of a verb, an object noun phrase and an adverb. The object noun phrase is the main constituent of the sentence while the adverb is usually an optional constituent. This grammatical difference between the sentence constituents might have been affected the interpretation of the sentence-final position as a focus cue. Several other studies (Gussenhoven, 1983a; Truckenbrodt & Darcy, 2010) have also noted that the difference between the main and optional constituents (sentence arguments *vs.* modifiers) affects sentence intonation. The second consideration is connected to the discussion in chapter 1.4 which states that Estonian can signal focus also with sentence-initial position and it has impressionistically reported to be a stronger cue for focus (Tael, 1988; Lindström, 2006).

Thus, a conclusion was drawn that the difference in grammatical function of constituents (argument *vs.* adverb) might have affected the sentence-final position as a focus cue and that the sentence-final position might not be such a strong cue for focus. These considerations lead to next studies, in which phonetics of focus was examined in speech production. The effect of sentence-initial position was investigated in a sentence consisting of main constituents (object and subject noun phrase) only.

6.2. Phonetics of sentence accent in Estonian

The first aim of the acoustic study of focus was to investigate whether accent shift that affected focus perception so effectively appears also in speech production. The participants were asked to utter sentences either with broad focus (as an answer to question *What happened?*) or with narrow focus on one of the sentence constituents (e.g. as an answer to question *Who drew a whale?*). The second aim of the study was to test the effect of sentence-initial position in OVS word order, where sentence-initial object noun phrase stands automatically in focus according to the theory of Estonian word order (Tael, 1988; Linström, 2006). In order to exhaust the strength of the sentence-initial position, OVS word order was embedded into context in which either broad focus or the focus on subject noun phrase was expected. Vainio and Järvikivi (2007) have carried out a similar study with the idea of exhausting the sentence-final position as a focus cue in Finnish. Their hypothesis was that the production of neutral

sentence intonation together with sentence-final focus position is impossible but their experiment showed that the speakers completely ignored the sentence-final position as focus and adjusted the sentence intonation to the focus implied by the context. These observations gave rise to two hypotheses. The first hypothesis relied on the results of previous perception study and predicted that the location of prosodic prominence in an utterance changes together with the focus of the sentence. The second hypothesis rests on the idea of overriding (Välimaa-Blum, 1988; 1993) and predicted that the intonation of OVS word order is adjusted to the pragmatic implications of the context (broad focus, focus on subject noun phrase).

The first hypothesis was substantiated. Speakers shifted the location of the accent for signalling focus of an utterance. The second hypothesis was partly confirmed. It was possible to adjust the intonation of OVS word order to the context with subject noun phrase in focus, but not to the context with the whole utterance in focus. In broad focus context, OVS word order was consistently produced with phrase-initial nuclear pitch accent. Thus, it was concluded that the 'default intonation' of OVS word order is phrase-initial nuclear pitch accent. However, if OVS word order is embedded into a context with narrow focus on the subject noun phrase, the default intonation can be replaced with phrase-final nuclear pitch accent. This, in turn, demonstrates that the pragmatic implication of the sentence-initial position can be overridden.

6.3. Phonetics of corrective focus

In connection to the acoustics of broad and narrow focus, also the acoustic difference between new information focus (e.g. *What did Lena draw? – Lena drew a whale.*) and corrective focus (e.g. *Lena drew a lion. – No! She drew a whale*) was investigated. The corrective focus being more 'emphatic' than new information focus has been quite frequently disputed in theoretical literature (Halliday, 1967a; Chafe, 1976; Rooth, 1992; Lambrecht, 1994). The hypothesis following some previous experimental studies (Breen et al., 2010; Hansen, 2011; Chen & Braun, 2006) predicted that corrective focus causes greater prosodic prominence than new information focus. The hypothesis did not gain support. In Estonian, corrective and new information focus do not differ in degree of prominence.

6.4. Phonetics of givenness

The perception study and the acoustic investigation of Estonian focus showed that in Estonian the accent shift is consistently used for focus marking and also perceived as a focus marker. Although, the language has free word order as an optional means for signalling focus and other information-structural categories. In this sense, Estonian sentence intonation appeared to function similarly to intonation in intonation languages. This outcome motivated the investigation of givenness (Chafe, 1976; Cruttenden, 2006; Baumann, 2006; Baumann & Riester, 2012; activeness in Lambrecht, 1994). Givenness causes deaccentuation (flattening of the F0 contour) in intonation languages (Brown, 1983; Terken & Hirschberg, 1994; Xu & Xu, 2005 for English; Swerts et al., 2002 for Dutch; Baumann, 2006 for German). The aim of the fourth and the last experimental study was to test whether given information in Estonian is deaccented.

Terken and Hirschberg (1994:138) have suggested in their study that the grammatical function and the sentence position might influence the realization of predicted deaccentuation. They propose that givenness preceding focus might not be realized by deaccentuation, but rather by pitch range variation. They did not find any evidence for the effect of sentence position or grammatical function for English. As one of the explanations they proposed for their result was that in English, there is a strong interaction between grammatical function and sentence position. As learned above, Estonian is a free word order language in which both word orders (SVO and OVS) are grammatical. Therefore, the hypothesis put forward in Terken and Hirschberg (1994) was tested for Estonian. The first hypothesis of the experiment predicted that the givenness preceding focus is realized with *prenuclear pitch accent* and the givenness following focus with *deaccentuation*. The second hypothesis predicted that the grammatical function has an effect on the degree of deaccentuation.

On the basis of the results the question whether Estonian deaccents on given expressions could be answered positively. The first hypothesis gained support. It appeared contrastively to Terken and Hirschberg (1994) that the prosodic effects of givenness depend on the sentence position: givenness preceding focus is signalled by a smaller pitch range of a prenuclear pitch accent, whereas at the end of an utterance following the focus givennes is signalled by the absence of pitch accent (deaccentuation). The second hypothesis did not gain support. There was no difference

between the object and subject noun phrase in the prosodic realization of givenness neither at the beginning of a sentence nor at the end of a sentence.

The conclusion of the study was that the structure of the intonation phrase (prenuclear *vs.* nuclear) affects the realisation of deaccentuation as grammatical means of givenness.

6.5. Disscussion and conclusion

The first question initiated in the introduction of this chapter was whether information structure has tonal correlates in Estonian. There are only a few studies on pitch prominence (Sahkai et al., 2013ab) in connection to information structure. Sentence accent has been a subject of impressionistic speculations (Erelt et al., 1993). The results of the perception experiment showed that the pitch accent in the nuclear position (as the last accent in a phrase) is interpreted as focus marker in free word order language Estonian. In speech production, focus was highlighted with pitch accent and givenness with F0 compression or deaccentuation.

Secondly, the thesis addressed the question whether information structure or word order interacts stronger with intonation in Estonian. Word order had a weak effect. Pitch accent and deaccentuation overrode the hypothetical pragmatic functions of sentence-final and sentence-initial position. Therefore, the hypothesis of overriding (Välimaa-Blum, 1988; 1993) got support. These results are compatible with studies of other free word order languages such as Finnish (Vainio and Järvikivi, 2007) or typologically unrelated Georgian (Skopeteas et al., 2009). What would explain this effect of overriding?

The introduction in chapter 1 discussed two information-structural categories other than focus that influence word order: identifiability and activeness. Recall from chapter 1.3 that Välimaa-Blum (1988) describes for Finnish that a constituent in the sentence-initial position refers to an identifiable referent, whereas a constituent in the sentence-final position to an unidentifiable referent. According to some studies (Tael, 1988; Erelt et al., 1993; Lindström, 2002; 2004), the case is similar for Estonian. The weak effect of sentence-final position in the perception experiment (chapter 2; section 6.1) might be evidence that encoding of identifiability is less important than focus for a spoken discourse.

The studies in chapters 3 and 4 dealt with the sentence-initial position. Similarly to the sentence-final position, a weak effect of word order occurred. OVS (object-verb-subject) word order that was hypothesized to mark the sentence-initial object noun phrase in focus was successfully produced with accent shift and accommodated with focus on subject noun phrase. As discussed in chapter 1.4, sentence-final position serves as focus position for the inverted subject. This probably explains why the speakers accented the subject noun phrase in the sentence-final position. The results on givenness in the fifth experiment can be explained along similar lines. The results are compatible with predictions proposed by Välimaa-Blum (1988). She predicts that word orders SVX (subject–verb–non-subject) and XVS (non-subject–verb-subject) are neutral[27] in terms of information structure. In Välimaa-Blum (1993) she provides data for Finnish that both SVX and XVS word orders can be produced with nuclear pitch accent on the object noun phrase or with nuclear pitch accent on subject noun phrase depending on the context.

As was discussed above, object-initial word order might also be connected to activation through poset relation (Ward & Birner, 2004; Weskott et al., 2011). In principle, a sentence-initial object noun phrase refers to the referent that is textually inferrable (e.g. the *bran muffin* that could be activated by *pastries*; Ward & Birner, 2004; Weskott et al., 2011). However, the referent activated by poset is semi-active. Baumann and Grice (2006) have found for German that semi-active referents are encoded with a particular type of pitch accent (HL*). Féry and Drenhaus (2008) have shown that the preferred position for the nuclear pitch accent in a sentence with OVS word order is in the beginning of a phrase. This preferred position of the nuclear pitch accent in OVS word order is consistent with the semi-active state of the initial sentence constituent. The results from Estonian presented in chapter 3 are compatible with the claim about German OVS. However, speech production experiments in chapters 3 and 4 also suggest that the semi-active state encoded by the sentence-initial position can be overridden by focus accent at the end of the intonational phrase.

Returning to the question of overriding, an explanation might be that word order is used for different information-structural categories that are possibly related to the information status of a referent within a discourse. Focus is a speaker's presentational

[27] She analyzes existential clause, but she implies that this analysis applies for all clause types in Finnish that enable free word order (Välimaa-Blum, 1988: 74).

choice and can override the information status of the discourse referents. The experiments conducted for the thesis concentrated on categories of information status – focus and givenness – that have strongly categorical effect on intonation: pitch accent *vs.* deaccentuation. The weak effect of word order might be explained by the possibility that the speakers were rather forced to ignore other discourse-pragmatic categories that might be important for word order.

We would like to draw attention to some of the methodological aspects of the study. In chapter 3 it was seen that the sentences in OVS word order were consistently uttered with a phrase-initial nuclear pitch accent in the broad focus context (as an answer to a question *What happened?*). However, this effect did not re-occur in the fifth experiment in which the production of OVS word order we examined in connection to givenness (chapter 5). Givenness was defined as a repetition of a noun phrase in a way like it shown in (6.1).

(6.1)

a. *Noored vahivad rannas.* 'Young people are hanging out on the beach.'
b. *Diivat kuulab muusik.* 'A musician listens to a diva.'
c. *Diivat kuulab Taavi.* 'Taavi listens to the diva.'

The noun phrase *diivat* ('diva') (6.1) occurs in (b) and is repeated in (c). Therefore, in (c) the *diivat* is given in the context of the utterances (a) and (c). The sentence in (a), in contrast, does not have any semantic relation to the sentence in (b). For this reason, (b) in context of the utterance in (a) is information-structurally neutral, also known as presentational or all-new focus (Gussenhoven, 2007). The comparison of the results presented in chapters 3 and 5 shows that the answer to the question and presentational focus are rendered with different accentuation patterns for OVS word order: with phrase-initial nuclear pitch accent in the former context and with phrase-final nuclear pitch accent in the latter. A tentative suggestion is that the pragmatics of a question, even of a general question such as *What happened?* may encourage the speech participant to focus on a particular referent. If nothing else is defined, then OVS word order elicits nuclear pitch accent on the object noun phrase.

In the methods, *F0 slope* related to time ('t Hart et al., 1990; Niebuhr, 2007) was calculated for the estimation of the accent or accent degree. The formula is shown in (6.2).

(6.2)

$$\text{Slope (st/s)} = (F0_{F01} - F0_{F02}) / (T_{F01} - T_{F02})$$

$F0_{F01}$ refers to F0 (st) of the first determining event in a pitch contour (F0 maximum or F0 minimum, elbow), $F0_{F02}$ refers to F0 (st) of the second determining event in a pitch contour (see methods section in chapter 2 for more details); T (ms) is the time of these events. The range of pitch excursion in semitones was divided by the duration of the pitch excursion. Slope with zero or close to zero means that there was no accent on the sentence constituent. Slope could provide acoustic basis for the typological studies similar to the study carried out by Cruttenden (2006). Cruttenden provides data on deaccentuation on a quite large sample of languages and reports the number of accents that was detected in the speech material. However, the process of accent determination was not described, but the perception of accent is kown to be strongly hearer-dependent.

Finally, the dissertational study left some questions open that could be finalized in future experiments. The idea that the grammatical function of the sentence constituent (argument *vs.* adverb) might have influenced the strength of sentence-final position as a focus cue was not pursued further. In addition, the acceptability of OVS word order with phrase-final nuclear pitch accent could be tested in a perception experiment. Different pragmatic contexts (pragmatic activation on object noun phrase *vs.* narrow focus on subject noun phrase) might show different effects in the acceptability of OVS word order.

In conclusion, the study provides a coherent overview for the intonational devices of information structure in Estonian. Focus is signalled with pitch accent and givenness is marked with F0-compression or deaccentuation. The effect of word order on sentence intonation was weak and the tonal correlates of focus and givenness overrode the information-structural implications that might have been encoded in the word order. Based on the theoretical considerations and results from the experiments, the thesis concludes that word order and intonation are sensitive to different aspects of information structure. Word order that might be encoding the information status of referents (or some other structures) is less important for intonation than a pitch accent that encodes the speaker's presentational choice – focus.

Zusammenfassung

Informationsstruktur beeinflusst die linguistische Übertragung von Information, neben anderen linguistischen und pragmatischen Prinzipien und mit der Sprachverarbeitung zusammenhängenden Faktoren. Kommunikation von neuer und relevanter Information setzt voraus, dass die alte, d.h. gegebene Information bei beiden Gesprächspartnern bekannt ist (Vallduví, 1992; Lambrecht, 1994). In vielen Fällen ist eine Äußerung so konstruiert, dass die Äußerung neuer Information auf alte/gegebene Information hinweist. Die Untersuchung der Informationsstruktur befasst sich mit der Frage, wie sich die Satzkomponenten auf den vorangegangenen Kontext und auf den Diskurs beziehen. Für die Diskussion der Informationsstruktur ist die Unterscheidung zwischen neuer und gegebener Information sehr relevant.

Nach Michael A. K. Halliday (1967ab) ist ein Sprachereignis in Informationseinheiten unterteilt. Die Informationseinheit ist gleichzeitig auch eine Intonationsphrase, d.h. phonologisch bedingt: „*one information unit is realized as one tone group*" (Halliday 1967a: 200). Pro Informationseinheit und gleichzeitig pro Intonationsphrase gibt es ein oder zwei Informationsschwerpunkte ("*point of information focus*", Halliday, 1967a: 204), die der Hervorhebungen in der Satzintonation (F0) entsprechen (*pitch prominence;* Halliday 1976a: 203). Dies impliziert, dass neue Information im Satz immer von prosodischer Prominenz begleitet sein muss.

Wallace Chafe (1974; 1976) verbindet die Informationsstruktur mit Bewusstsein und dem Wissen der Gesprächspartner: gegebene Information ist das Wissen, von dem der Sprecher denkt, dass der Hörer es während des Sprechaktes besitzt (Chafe, 1976: 30). Diese Definition setzt voraus, dass der Sprecher verpflichtet und fähig ist, Rücksicht auf das Wissen des Hörers zu nehmen. Ähnlich zu Halliday (1967ab), verbindet Chafe (1976, 1987) den Informationsstatus mit der Satzprosodie: "*given information is conveyed in a weaker and more attenuated manner than new information*" (Chafe 1976: 31). Chafe (1974; 1976; 1987) beschreibt Gegebensein (*givenness*, Chafe 1967: 31) eher im Sinne von Aktivierung und die grammatische Kategorie Bestimmtheit eher im Sinne von Identifizierbarkeit (Chafe, 1967: 39). Diese Konzepte hat Lambrecht (1994) auch als Aktivierung und Identifizierbarkeit

übernommen. Knud Lambrecht (1994) übernimmt weitestgehend die theoretischen Vorgehensweisen von Halliday (1967a) und Chafe (1974; 1976; 1987) und beschreibt Informationsstruktur in vier Kategorien, die für die linguistische Übertragung der Information relevant sind: Aktivierung (*activeness*), Identifizierbarkeit (*identifiability*), Thema (*topic*) und Fokus (*focus*).

Für die vorliegende Studie besonders wichtig ist Lambrechts Erkenntnis (1994: 208), die sich von Halliday (1967a) und Chafe (1976) unterscheidet, dass der Fokus getrennt von Satzakzentuierung behandelt werden muss. Lambrecht (1994) definiert die Kategorien der Informationsstruktur ausgehend von ihren Funktionen im Diskurs und unabhängig von der linguistischen Form. Linguistische Mittel, wie z.B. satzinitiale Position, Satzakzentuierung, Deakzentuierung und Wortstellung, sind grammatische Mittel, die beliebig für den Ausdruck der Diskursfunktionen eingesetzt werden können (Lambrecht 1994). Diese Dissertation folgt Lambrechts (1994) Definitionen von linguistischen Kategorien der Informationsstruktur, weil dieser Zugang explizit die Notwendigkeit vorsieht, dass die Kategorien der Informationsstruktur separat von linguistischen Mitteln definiert werden muss. Dieses Vorgehen eignet sich für die gleichzeitige Untersuchung von prosodischen und syntaktischen Strukturen.

Aus diesem Grund verwendet die vorliegende Arbeit die Kategorien Fokus, Aktivierung und Identifizierbarkeit, wie von Lambrecht (1994) definiert. Der **Fokus** ist der Teil des Satzes, den der Sprecher als relevante oder neue Information präsentiert (Halliday, 1967; Lambrecht, 1994). Aktivierung und Identifizierbarkeit bezeichnen pragmatischen Eigenschaften der Satzkomponenten in der Kenntnis der Gesprächspartner (Chafe, 1976; Lambrecht, 1994). **Aktivierung** bezieht sich auf den Status des Referenten im Bewusstsein der Gesprächspartner. Ein Referent kann durch Erwähnung, aber auch durch inhaltliche oder situative Rückschlüsse aktiviert werden. **Identifizierbarkeit** bezieht sich auf das Wissen der Gesprächspartner über den Referent. Die Satzkomponenten, die sich auf Referenten beziehen, die der Hörer schon kennt, sind für ihn identifizierbar. Wenn z.B. der Sprecher einen Hund hat und der Hörer den Hund kennt, dann ist der Hund als Referent im Satz "*Ich bin zu spät gekommen,weil ich noch den Hund gefüttert habe"* vom Hörer identifizierbar. Lambrecht (1994) definiert die Aktivierung und Identifizierbarkeit als binär (aktiviert *vs.* nicht-aktiviert, identifizierbar *vs.* nicht-identifizierbar). Für Chafe (1976) ist **Gegebensein** (*givenness*) allerdings ein Kontinuum. Stefan Baumann und seine

Kollegen (Baumann, 2006; Baumann und Grice, 2006; Baumann und Riester, 2012) halten am Begriff Gegebensein fest und zeigen, dass in einer Äußerung die verschiedenen Stufen der Aktivierung in der Satzintonation erkennbar sind: Die gegebene Information ist deakzentuiert, die aus dem Kontext ableitbare Information ist mit einem frühen Gipfel (HL*) und die nicht vorhersagbare Information mit einem mittleren (H*) Gipfel im Deutschen kodiert (Baumann und Grice, 2006: 1655). Im experimentellen Teil dieser Arbeit wird auch der Begriff Gegebensein bevorzugt.

Lambrecht (1994) diskutiert, getrennt für jede Kategorie und für mehrere Sprachen, die linguistischen Mittel, die von Sprechern zum Ausdruck der Kategorien genutzt werden können. Beispiele für Fokus sind Satzakzentuierung oder Wortstellung (Lambrecht 1994: 225, 319), für Aktivierung Deakzentuierung, Pronomina oder satzinitiale Position (Lambrecht 1994: 107) und für Identifizierbarkeit lexikalische Mittel (Pronomen *vs.* vollständige Substantiv (*full noun*), bestimmte *vs.* unbestimmte Artikel, Lambrecht 1994: 79). Diese Zuordnungen sind nicht bindend: z.B. ist ein Pronomen nicht automatisch Aktivierung oder ein Tonakzent Fokus.

Die vorliegende Studie untersucht in erster Linie den Tonakzent und die Wortstellung als grammatische Mittel der Informationsstruktur. In der theoretischen Einleitung im Kapitel 1 wird dargestellt, dass in mehreren Sprachen, die als Intonationssprachen (Englisch, Deutsch, Niederländisch) gelten, Tonakzent und Akzentverschiebung als primäre Merkmale (*cues*) für Fokus im Satz verwendet werden, während Deakzentuierung auf Aktivierung/gegebene Information hinweist (Halliday, 1967b; Ladd, 2008; Brown, 1983; Cruttenden, 2006; Krahmer und Swerts, 2001; Breen et al., 2010; Kohler, 1991; Féry und Kügler, 2008; Baumann et al., 2006; Baumann und Grice, 2006). In einigen der nicht-konfigurativen Sprachen (Ungarisch, Finnisch, Italienisch, Griechisch, Georgisch) wurde gefunden, dass der Fokus durch die Wortstellung (z.B. in der präverbalen Position) signalisiert werden kann (Kiss, 1995; Rizzi, 1997; Keller und Alexopoulou, 2001; Skopeteas et al., 2009). Diese zwei linguistischen Mittel, (De-)Akzentuierung und Wortstellung, wurden häufig als - innerhalb derselben Sprache - sich gegenseitig ausschließend behandelt. Lambrecht (1994: 240) schlägt sogar vor, dass Wortstellung die gleiche pragmatische Funktion wie Tonakzente erfüllt. Darüber hinaus diskutiert die theoretische Einleitung der Arbeit, ob Wortstellung für andere unterschiedliche pragmatische Funktionen eingesetzt wird als der Tonakzent. Abschließend wird festgestellt, dass sowohl der Tonakzent und

Wortstellung für mehrere pragmatische Funktionen geeignet sind, als auch sehr häufig in einer Sprache gleichzeitig vorkommen.

Im experimentellen Teil der Arbeit wird das Estnische zur Erforschung vom Zusammenhang zwischen grammatischen Mitteln und Fokus sowie Gegebensein genutzt. Estnisch gehört innerhalb der Finnougrischen Sprachen zur finnischen Sprachgruppe (Abondolo, 1998; Viitso, 2003) und steht daher der finnischen Sprache näher als dem Ungarischen. Estnisch ist eine der Sprachen, die eine relativ freie Wortfolge im Satz erlaubt und dafür bekannt ist, dass die Wortfolge verschiedene Aspekte der Informationsstruktur (Fokus, Bestimmtheit) markiert (Tael, 1988; Erelt et al, 1993; Vilkuna, 1998; Lindström, 2002; 2004). Das Intonationssystem des Estnischen ist sehr vielfältig (Asu, 2004; 2005) und es ist unklar, ob z.B. der Fokus auch mit Hilfe der Intonation (*pitch prominence*) vermittelt wird. Impressionistische Beobachtungen enthalten Hinweise, dass Satzakzentuierung für die Äußerung des Fokus benutzt wird (Erelt et al., 1993). In den Kapiteln 2, 3, 4 und 5 werden die prosodischen Effekte von Fokus und Gegebensein im Zusammenhang mit der Wortstellung experimentell untersucht:

Interagiert Satzintonation stärker mit Informationsstruktur oder mit Wortstellung? In den folgenden fünf Abschnitten werden die Ergebnisse der Experimente zusammengefasst und anhand der möglichen theoretischen Perspektiven diskutiert.

Wahrnehmung der Satzakzentuierung und Akzentverschiebung (*accent shift*)

Intonationssprachen, wie das Englische oder Deutsche, zeigen eine feste Beziehung zwischen dem Tonakzent und dem Fokus (z.B. Ladd, 2008; Halliday 1967b). Zusammen mit der Akzentverschiebung verändert sich auch der Fokus des Satzes (Swerts et al., 2001; Breen et al., 2010). Genauso zeigen impressionistische Beobachtungen des Estnischen (Erelt et al., 1993), dass sich der Fokus des Satzes durch Akzentverschiebung verändert. Die Theorie der estnischen Wortfolge gibt an, dass die satzfinale Position eine Position für Fokus mit neuer Information ist (Erelt et al., 1993; Lindström, 2006). Dies bedeutet, dass satzfinale Satzglieder automatisch als Fokus des Satzes interpretiert werden. Vainio und Järvikivi (2006) zeigen anhand eines Versuchs, dessen Vorgehen von dem hier dargestellten Experiment leicht abweicht, dass im Finnischen die satzfinale Position, beim Auftreten eines bestimmten Typs von

Satzgliedern (Lokaladverbialien), ein starker Merkmal (*cue*) für Fokus ist. Ähnlich dieser Studie zum Finnischen (Vainio und Järvikivi, 2006) befasst sich das hier dargestellte Experiment mit der Frage: was nehmen estnische Muttersprachler als Fokus wahr: Tonakzent oder satzfinale Position. Die Teilnehmer wurden gebeten auditiv präsentierte Äußerungen mit verschiedenen Wortstellungen und Akzentverlagerungen anzuhören und Kontexten mit verschiedenen Foki zuzuordnen.

Die Auswertung dieses Perzeptionsexperiments zeigte, dass die satzfinale Position keinen Einfluss auf die Fokus-Wahrnehmung im Estnischen hatte. Der nukleare Tonakzent wurde als Fokus des Satzes wahrgenommen und die Akzentverschiebung verursachte eine Veränderung in der Interpretation der Äußerung. Diese Ergebnisse führten zu zwei Überlegungen. Erstens: das Sprachmaterial in diesem Experiment unterschied sich vom Sprachmaterial im Experiment von Vainio und Järvikivi (2006). Deren Teilnehmer hörten Äußerungen, die aus einem Verb und zwei Satzadverbialien bestanden. Die Äußerungen im vorliegenden Experiment bestanden aus Verb, Objekt und Adverbial. Es ist möglich, dass die satzfinale Position mit der grammatischen Funktion interagierte. Auch weitere Intonationsstudien (Gussenhoven, 1983a; Truckenbrodt und Darcy, 2010) haben Hinweise gefunden, dass der Unterschied zwischen obligatorischen und optionalen (Satzargumente *vs.* Satzattribute) Satzgliedern die Satzintonation beeinflusst. Zweitens: Studien zur estnischen Wortstellung haben festgestellt, dass die satzinitiale Position als Merkmal für Fokus stärker und ‚emphatischer' als die satzfinale Position ist (Tael, 1988; Lindström, 2006).

Die Ergebnisse des dargestellten Experiments zeigen, dass die satzfinale Position als Merkmal für Fokus im Estnischen nicht stark genug ist. Aufbauend auf dieses Ergebnis wurde im zweiten Experiment der Einfluss der satzinitialen Position auf die Satzintonation in Äußerungen mit Hauptkonstituente (Nominalphrasen als Subjekt und Objekt) in einem Sprachproduktionsexperiment untersucht.

Phonetik der Satzakzentuierung im Estnischen

Die Ergebnisse des Perzeptionsexperiment zeigen, dass die Akzentverschiebung eine Änderung im Fokus der Äußerung verursacht. Die zweite Studie sollte klären, ob die Akzentverschiebung auch in der Sprachproduktion angewendet wird. Die Teilnehmer wurden gebeten Sätze entweder mit breitem Fokus (als Antwort auf die Frage *Was ist passiert?*) oder mit einem Satzglied im engen Fokus (z.B. als Antwort auf die Frage

Wer hat den Wal gemalt?) zu produzieren. Nach der Theorie der estnischen Wortfolge (Tael, 1988; Lindström, 2006) sollte das satzinitiale Objekt automatisch im Fokus stehen und es ist zu erwarten, dass die satzinitiale Position ein stärkeres Merkmal für Fokus ist als die satzfinale Position. Um die satzinitiale Position als Merkmal für Fokus auszuschöpfen, wurde der Satz mit der Wortstellung OVS (Object-Verb-Subjekt) in zwei verschiedene Kontexten gebettet: zum einen in einen Kontext mit breitem Fokus und zum anderen in einen Kontext mit dem Subjekt im engen Fokus. Vainio und Järvikivi (2007) haben ein ähnliches Experiment zur satzfinalen Position im Finnischen durchgeführt. Ihre Hypothese war, dass es nicht möglich ist einen Satz zu produzieren, dessen eines Satzglied durch satzfinale Position im Fokus steht und gleichzeitig mit neutraler Satzintonation, wie sie bei breitem Fokus gefunden wird, realisiert ist. Dennoch haben ihre Sprecher die satzfinale Position ignoriert und die Satzakzentuierung an die pragmatischen Implikationen angepasst, die der Kontext vorausgesetzt hat. Von der Theorie ausgehend wird in Bezug zur satzinitialen Position im Estnischen erwartet, dass die neutrale Intonation des breiten Fokus nicht möglich ist.

Die erste Hypothese, dass die Akzentverschiebung auch in der Sprachproduktion auftritt, wurde bestätigt. Die zweite Hypothese wurde zum Teil bestätigt. Zwar konnte die Satzakzentuierung der Äußerungen mit der Wortfolge OVS und dem Subjekt im engen Fokus angepasst werden aber nicht bei breitem Fokus. Im breiten Fokus wurde die Wortfolge OVS immer mit satzinitialem Nuklearakzent produziert. Daraus ist ableitbar, dass es nicht möglich ist, den Satz, der ein Satzglied durch satzinitiale Position im Fokus hat, mit neutraler Intonation des breiten Fokus zu produzieren. Die Wortfolge OVS konnte jedoch mit einem satzfinalen Nuklearakzent im Kontext mit engem Fokus auf dem Subjekt produziert werden. Die Schlussfolgerung daraus ist, dass die sogenannte ‚*default*'-Intonation von der Wortfolge OVS im engen Fokus ersetzt werden könnte.

Dieses Ergebnis spricht für die Hypothese der *overriding,* wie von Välimaa-Blum (1988, 1993) vorgestellt. Diese Hypothese besagt, dass im Finnischen (aber vielleicht auch in anderen Sprachen mit freier Wortfolge) die verschiedenen Wortstellungen als vollständige Konstruktionen (Fillmore, 1987) mit eigenen semantischen, pragmatischen und auch prosodischen Eigenschaften fungieren. Wenn die Konstruktion aber in einem Kontext auftritt, dessen pragmatische Implikationen von den Implikationen der Konstruktion abweichen, ist es möglich die

konstruktionsbedingten pragmatischen Implikationen durch die Satzakzentuierung zu ersetzen.

Phonetik des korrektiven Fokus im Estnischen

Im Zusammenhang mit dem akustischen Unterschied zwischen breitem und engem Fokus wurde auch die Produktion des korrektiven Fokus untersucht. Ob der korrektive Fokus (z.B. „Lena malte die Löwe." – „Nein! Sie malte den Wal") ‚emphatischer' ist als der Fokus für neue Information (z.B. „Was malte Lena?" – „Lena malte den Wal."), wird in der theoretischen Literatur (Halliday, 1967a; Chafe, 1976; Rooth, 1992; Lambrecht, 1994) oft disputiert. Basierend auf experimentellen Studien (Breen et al., 2010; Hanssen, 2008; Chen und Braun, 2006), wurde die Hypothese getestet, ob der korrektive Fokus im Estnischen prosodisch mehr Prominenz verursacht. Es fanden sich dafür keine Hinweise: Im Estnischen unterscheiden die zwei Foki in der prosodischen Prominenz nicht.

Phonetik des Gegebenseins (*givenness*)

Die Perzeptions- und Sprachproduktionsexperimente zeigten überzeugend, dass der Tonakzent im Estnischen als Fokus wahrgenommen wird und dass die Sprecher den Fokus im Satz mit Tonakzent produzieren. Auf diese Weise funktioniert Intonation im Estnischen ganz ähnlich wie in Intonationssprachen. Für Intonationssprachen ist bekannt, dass die Sprachausdrücke, die im Diskurs aktiviert sind, d. h. gegebene Information beinhalten, deakzentuiert sind (Chafe, 1976; Brown, 1983; Cruttenden, 2006; Baumann, 2006; Baumann und Riester, 2012). Das erste Ziel des vierten Experiments war zu testen, ob die gegebene Information auch im Estnischen deakzentuiert wird. Die weiteren Forschungsziele sind von Terken und Hirschberg (1994) motiviert, die untersuchten, inwieweit Position im Satz und grammatische Funktion (z.B. Subjekt *vs.* Objekt) Deakzentuierung beeinflussen. Terken und Hirschberg (1994) nahmen an, dass Gegebensein (*givenness)* in der Position nach dem Fokus mit Deakzentuierung, vor dem Fokus aber mit der Variation im Tonumfang (*pitch range variation*) produziert wird. Daher ist die Erwartung, dass die gegebene Information vor dem Fokus mit einem *prenuklearen Tonakzent* und nach dem Fokus mit D*eakzentuierung* produziert wird. Terken und Hirschberg (1994) fanden keinen starken Beweis, dass die grammatische Funktion oder Satzposition Deakzentuierung

beeinflusst. Sie schlagen vor, dass einer der Gründe, der die Ergebnisse beeinflusst haben könnte, die starke Interaktion zwischen der Satzposition und der grammatischen Funktion im Englischen ist. Wie oben gesehen, sind im Estnischen beide Wortfolgen SVO und OVS grammatikalisch, was es ermöglicht, die Effekte von grammatischer Funktion (Subjekt *vs.* Objekt) auf Deakzentuierung des Gegebenseins unabhängig von der Satzposition zu testen.

Die erste Hypothese wurde bestätigt. Im Estnischen wird gegebene Information deakzentuiert. Die zweite Hypothese wurde auch bestätigt: in prenuklearer Position trägt der Sprachausdruck mit gegebener Information einen prenuklearen Tonakzent während in postnuklearer Position die gegebene Information deakzentuiert ist. Die dritte Erwartung hat keine Unterstützung gefunden: die prosodische Realisation von Gegebensein unterschied sich zwischen Subjektphrase und Objektphrase nicht.

Diskussion und Schlussfolgerung

Wie oben genannt, lautete die Forschungsfrage: Was interagiert stärker mit der Satzintonation: Wortstellung oder Informationsstruktur? Die Ergebnisse zeigen, dass die Wortstellung wenig Einfluss auf die Satzintonation hat. In den Perzeptions- und Produktionsexperimenten wurde gezeigt, dass der Fokus mittels Tonakzent vermittelt wird, während das Gegebensein eine klare F0-Komprimierung oder Deakzentuierung verursacht. In der theoretischen Perspektive wurde gleichzeitig festgestellt, dass die hypothetischen pragmatischen Funktionen der satzfinalen oder satzinitialen Position durch die prosodischen Mittel der Informationsstruktur aufgehoben werden. Eine Möglichkeit, dies zu erklären, ist, dass Wortstellung und prosodische Mittel im Estnischen für verschiedene pragmatische Funktionen eingesetzt werden.

In der Einleitung der Arbeit (Kapitel 1) wurde bereits erwähnt, dass die Funktionen, die die verschiedenen Wortstellungen signalisieren können, Identifizierbarkeit und Aktivierung sein könnten. Riitta M. Välimaa-Blum (1988, und die Referenzen dort) gibt an, dass sich im Finnischen der Sprachausdruck, der am Satzanfang oder vor dem Verb steht, auf einen Referent, der für den Hörer identifizierbar und lokalisierbar ist (wie *der Hund* in dem obengenannten Beispiel) bezieht. Nach den wenigen veröffentlichten Untersuchungen (Tael, 1988; Lindström, 2004) könnte dies im Estnischen ähnlich sein. Der Theorieansatz in Lambrecht (1994) würde voraussagen, dass sowohl identifizierbare als auch nicht-identifizierbare

Referenten sich im Fokus befinden können. Das letzte Satz im Beispiel (1.3) im Kapitel 1 stellt eine Äußerung dar, in dem der Referent, der schon im Diskurs erwähnt wurde, im Fokus steht. Im Fall des Estnischen, wo die satzinitiale und satzfinale Position von der Akzentuierung dominiert wird, könnte als analog gelten, da die Identifizierbarkeit für die Sprecher weniger wichtig als der Fokus ist.

In der Einleitung vom Kapitel 1 wird die Variation der Wortstellung, die unter bestimmten pragmatischen Bedingungen auch im Englischen und im Deutschen möglich ist, umfassend beschrieben. Das sogenannte *preposing* (Ward und Birner, 2004), umfasst die objekt-initiale Wortstellungen wie OSV im Englischen (Ward und Birner, 2004) und OVS im Deutschen (Weskott, et al., 2011). Auffällig an diesen objekt-initialen Wortstellungen ist, dass sie in Kontexten gelingen, in denen sich die Objektphrase auf den Referent beziehen, der aus dem Kontext ableitbar ist (z.B. *die Brezel* in der Äußerung *Ich habe das ganze Gebäck für das Frühstück gekauft, die Brezel habe ich aber vergessen*). Die satzinitiale Position im Estnischen könnte eine ähnliche Art von Aktivierung der Referenz markieren. Die Ergebnisse der oben beschriebenen Experimente könnten darauf hinweisen, dass im Estnischen auch diese Art der Aktivierung von der pragmatischen Funktion des Fokus aufgehoben werden kann.

Zusammenfassend geben die experimentellen Studien und die theoretische Analyse dieser Doktorarbeit Hinweise darauf, dass die verschiedenen Wortstellungen und die prosodischen Mittel im Estnischen unterschiedlichen Aspekten der Informationsstruktur dienen. Weitere Untersuchung ist notwendig, aber die vorläufige Folgerung ist, dass der Fokus definitiv mit der prosodischen Prominenz ausgedrückt wird, während die Identifizierbarkeit oder Aktivierung in der Wortstellung kodiert sein können. Dennoch muss auch zugelassen werden, dass auch die Aktivierung/Gegebensein durch Prosodie, nämlich Deakzentuierung, übermittelt wird. Daher hat die hypothetisierte Aktivierung, markiert durch satzinitiale Position, vermutlich noch zusätzliche pragmatische Eigenschaften. Generell zeigen die Daten aus dem Finnischen (Välimaa-Blum, 1988, 1993; Vainio und Järvikivi, 2006; 2007) und dem Estnischen (diese Studie), dass die linguistische Szene der grammatischen Mittel der Informationsstruktur deutlich vielfältiger ist, als es die Dichotomie zwischen Tonakzentsprachen und Sprachen mit freier Wortfolge hervorsagen würde (ähnliche Schlussfolgerung sind auch in Face und D'Imperio (2005) zu finden). Die theoretischen

Überlegungen und experimentellen Versuche weisen darauf hin, dass eine Sprache über mehrere sprachliche Mittel gleichzeitig verfügt, sowohl Tonakzent als auch verschiedene Wortstellungen, und diese für verschiedene pragmatische Funktionen einsetzt.

Bibliography

Aare, K., Lippus, P. & Šimko, J. (2014). Creaky voice in spontaneous spoken Estonian. Jähi, K. & Taimi, L. (Eds.), *Proceedings of XXVIIIth Phonetics Symposium "Fonetiikan päivät"*. Finland, Turku 2013 (pp. 27-35). Turku: Turun yliopisto

Abbott, B. (2006). Definiteness and indefiniteness. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 122-149). Blackwell Publishing Ltd.

Abondolo, D. (1998). Introduction. In Abondolo, D. (Ed.) *The Uralic Languages*. New York: Routledge

Arumäe, T. (2014). Autosegmenteerija [Computer program]. Retrieved in 2014 from < https://phon.ioc.ee/dokuwiki/doku.php?id=projects:tuvastus:est-align.et>.

Ariste, (1939). A Quantitative Language. – *Proceedings of the Third International Congress of Phonetic Sciences* (pp. 276–280). Ghent.

Arnhold, A. (2014). Finnish prosody: studies in intonation and phrasing. Doctoral thesis, Johann Wolfgang Goethe-Universität zu Frankfurt am Main.

Arvaniti, A., Ladd, D.R., & Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*, 26, 3-25.

Arvaniti, A., Ladd, D.R., & Mennen, I. (2006). Phonetic effects of focus and 'tonal crowding' in intonation: evidence from Greek polar questions. *Speech Communication*, 48, 667-396.

Asu, E. L. (2004). The phonetics and phonology of Estonian intonation. Doctoral dissertation, University of Cambridge.

Asu, E. L. (2005). Towards a phonological model of Estonian intonation. *Proceedings of the Second Baltic Conference on Human Language Technologies*, Tallinn 4-5 May 2005 (pp. 95-100).

Asu, E. L. (2006). Rising intonation in Estonian: an analysis of map task dialogues and spontaneous conversations. R. Aulanko, L. Wahlberg & M., Vainio (Eds.), *Fonetiikan Päivät 2006 – The Phonetics Symposium 2006* (pp. 1-8). Helsinki: Helsinki University.

Asu, E.L., & Nolan, F. (1999). The effect of intonation on pitch cues to the Estonian quantity contrast. In Ohala, J.J. (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco 1-7 August 1999 (pp. 1873-1876). University of California

Asu, E. L. & Nolan, F. (2006). Estonian and English rhythm: a two-dimensional quantification based on syllables and feet. In R. Hoffmann & H. Mixdorff (Eds.), *Speech Prosody – Proceedings of the third International Conference*, Dresden, 2-5 May 2006 (pp. 249-252). Dresden: TUDpress

Asu, E. L. & Nolan, F. (2007). The Analysis of Low Accentuation in Estonian. *Language and Speech*, 50(4), 567-588.

Asu, E.L. & Teras, P. (2009). Illustrations of the IPA: Estonian. *Journal of the International Phonetic Association*, 39(3), 367-372.

Asu, E.L, Lippus, P., Teras, P. & Tuisk, T. (2009). The realization of Estonian quantity characteristics in spontaneous speech. In M. Vainio, R. Aulanko & O. Aaltonen (Eds.), *Nordic Prosody – Proceedings of the Xth Conference*, Helsinki 2008 (pp. 49-56). Frankfurt: Peter Lang Verlag

Atterer, M., & Ladd, D.R. (2004). On the phonetics and phonology of segmental anchoring of f0: evidence from German. *Journal of Phonetics*, 32, 177-197.

Ayers, G. (1996). Nuclear accent types and prominence: some psycholinguistic experiments. Doctoral dissertation, The Ohio State University, Columbus.

Baker M.C., (2001). Configurality and polysynthesis. In M. Haspelmath, E. König, W. Oesterreicher & W. Raible (Eds.), *Language typology and language universals: an international handbook*, (Vol. 2, pp. 1433-1441). Berlin: Walter de Gruyter.

Bard, E., Sotillo, C., Anderson, A.H., Doherty-Sneddon, G & Newlands, A. (1995). The control of duration in running speech. In K. Elenius and P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol.4). Stockholm: Stockholm University.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.

Baumann, S. (2006). The Intonation of Givenness - Evidence from German. Doctoral dissertation, Saarland University. Linguistische Arbeiten 508. Tübingen: Niemeyer

Baumann, S. & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics*, 38, 1636-1657.

Baumann, S. & Hadelich, K. (2003). Accent type and givenness: an experiment with auditory and visual priming. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain (pp.1811-1814).

Baumann, S. & Riester, A. (2012). Referential and lexical givenness: semantic, prosodic and cognitive aspects. In G. Elordieta & P. Prieto (Eds.), *Prosody and Meaning. Interface Explorations 25*. Berlin, New York: Mouton De Gruyter. 119-162.

Baumann, S., Grice, M. & Steindamm, S. (2006). Prosodic Marking of Focus Domains - Categorical or Gradient? *Speech Prosody – Proceedings of the third International Conference*, Dresden, 2-5 May 2006 (pp. 301–304). Dresden: TUDpress Verlag der Wissenschaften.

Beckman, M. E. (1986). Stress and non-stress accent (Netherlands Phonetics Archives, vol. VII). Dordrecht: Foris Publication Holland.

Beckman, M. E. & Elam, G. A. (1997). Guidelines for ToBI Labelling. The Ohio State University Research Foundation. Retrieved on 20.02.2015 from http://www.cs.columbia.edu/~agus/tobi/labelling_guide_v3.pdf

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P.A. Keating, (Eds.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III* (pp. 7-33). Cambridge University Press.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255-309.

Boersma, P., & Weenink, D. (2012). Praat: Doing phonetics by computer [Computer program]. Retrieved in 2012 from <http://www.praat.org/>.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer [Computer program]. Rertrived in 2013 from <http://www.praat.org/>.

Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer [Computer program]. Rertrived in 2014 from <http://www.praat.org/>.

Breen, M., Fedorenko, E., Wagner, M., Gibson, E. (2010). Acoustic correlates of information structure. *Language and cognitive processes*, 25: 7, 1044–1098.

Brown, (1983). Prosodic Structure and the Given/New Distinction. In A. Cutler & L. R. Ladd (Eds.), *Prosody: Models and Measurements.* Springer Series in Language and Communication, Vol. 14, 67-77.

Bruce, G. (1977). Swedish word accents in sentence perspective. Lund: Geerups.

Büring, D. (2009). Towards a typology of focus realization. In M. Zimmermann & C. Féry (Eds.), *Information Structure. Theoretical, Typological, and Experimental Perspectives* (pp. 177–205). Oxford: Oxford University Press.

Cambier-Langeveld, T. & Turk, A. (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics* 27, 255-280.

Campbell, Nick & Mary Beckman (1997). Stress, Prominence, and Spectral Tilt. *Intonation: Theory, Models and Applications – Proceedings of an ESCA Workshop*, Athens, Greece, 1998 (pp. 67-70).

Chafe, W. L. (1974). Language and consciousness. *Language* 50, 111-133.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. Li (Ed.), *Subject and Topic* (pp. 67-78). London: Academic Press, Inc.

Chafe, W. (1987). Cognitive constraints on information flow. In R. S. Tomlin (Ed.), *Coherence and Grounding in Discourse: Outcome of a Symposium*, Eugene, Oregon, June 1984 (pp. 21–51).

Chen, Y. & Braun, B. (2006). Prosodic realization of information structure categories in Standard Chinese. *Speech Prosody – Proceedings of the third International Conference*, Dresden, 2-5 May 2006. Dresden: TUDPress

Chomsky, N. (1957). Syntactic structures. The Hague/Paris: Mouton

Clark, H. H. (1996). Using language. Cambridge: Cambridge University Press

Clark, H. H. & Haviland, S. E. (1977). Comprehension and the given-new contract. In Freedle, R. O. (Ed.), *Discourse processes: advances in research and theory* (Vol. 1, 1-40). Norwood, New Jersey: Ablex Publishing Corporation.

Cohen, A., Collier, R. & 't Hart J.(1982). Declination: construct or intrinsic feature of speech pitch? *Phonetica* 39, 254-273.

Collier, R. (1974). Laryngeal muscle activity, subglottal air pressure, and the control of pitch in speech. Status report on speech research SR-39/40. Haskins Laboratories-

Cooper, W., Eady, S. J., Mueller P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Journal of Acoustical Society of America*, 77, 2142–55.

Couper-Kuhlen, E. (1984). A new look at contrastive intonation. In R. J. Watts & U. Weidmann, (Eds), *Modes of Interpretation: Essays Presented to Ernst Leisi* (pp. 137–158). Tubingen: Gunter Narr Verlag.

Cruttenden, A. 2006. The De-accenting of Given Information: a Cognitive Universal? In Bernini, G. & M. L. Schwartz (Eds.), *Pragmatic Organization of Discourse in the Language of Europe* (pp. 311-355). The Hague: Mouton de Gruyter.

Cutler, A. & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, 20, 1-10.

D'Imperio, (2000). The role of perception in defining tonal targets and their alignment. Doctoral dissertation, The Ohio State University.

Daneš, F. (1966). A three-level approach to syntax. In Daneš et al. (Eds.) *Travaux linguistiques de Prague* (Vol. 1, 225-240). University of Alabama Press.

Draxler, Chr. (2011) Percy - an HTML5 framework for media rich web experiments on mobile devices; *Proceedings of Interspeech Florence*, Italy (pp. 3339-3340).

É. Kiss, K. (1995). Introduction. In K. É. Kiss (Ed.) *Discourse configurational languages. Oxford studies in comparative syntax* (pp. 3–27). New York, Oxford: Oxford University Press.

É. Kiss, K. (2001). Discourse configurationality. In M. Haspelmath, E. König, W. Oesterreicher & Raible, W. (Eds.), *Language typology and language universals: an international handbook*, (Vol. 2, pp. 1442-1455). Berlin: Walter de Gruyter.

É. Kiss, K. (2004). The Syntax of Hungarian. Cambridge: Cambridge University Press

Eady, J. S. & Cooper, E. W. (1986). Speech intonation and focus location in matched statements and questions. *Journal of Acoustical Society of America*, 80(2), 402-415.

Eek, A. & Meister, E., (1997). Simple perception experiments on Estonian word prosody: foot structure vs. segmental quantity. In I. Lehiste & J. Ross, (Eds.), *Estonian prosody: Papers from a symposium* (pp. 71–99). Tallinn: Institute of Estonian Language,.

Eek, A., & Meister, E. (2003). Foneetilisi katseid ja arutlusi kvantiteedi alalt (I). Häälikukestusi muutvad kontekstid ja välde. *Keel ja Kirjandus*, 11, 815–837; 12, 904–918.

Eek, A., & Meister, E. (2004). Foneetilisi katseid ja arutlusi kvantiteedi alalt (II). Takt, silp ja välde. *Keel ja Kirjandus*, nr 4, 251–271; 5, 336–351.

Eek, A. (1974). Observations on the duration of some word structures I. *Estonian Papers in Phonetics* (pp. 31–57).

Eek, A., & Help, T. (1987). The interrelationship between phonological and phonetic sound change: a great rhythm shift of Old Estonian. *Proceedings of the 11th International Congress of Phonetics Sciences*, Estonia, Tallinn 1-7 August 1987.

Eek, A. (1980). Estonian quantity: notes on the perception of duration. In A. Eek (Ed.), *Estonian Papers in Phonetics* (pp. 5–30). Tallinn: Academy of Sciences of the Estonian S.S.R.

Eek, A. (1983). Kvantiteet ja rõhk eesti keeles (I). *Keel ja Kirjandus*, 9, 481– 489.

Eek, A., (1990). Units of temporal organisation and word accents in Estonian. *Linguistica Uralica*, 26(4), 251–263.

Ehala, M. (2006). The word order of Estonian: Implications to Universal Language. *Journal of Universal Language*, 7, 49-89.

Erelt, M. & Metslang, H. (2006). Estonian clause patterns - from Finno-Ugric to Standard Average European. *Linguistica Uralica*, XLII, 4, 254-266.

Erelt et al., (1993). Eesti keele grammatika. Süntaks. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

Erelt, M. (2003). Preface. *Estonian Language. Linguistica Uralica. Supplementary series,* Vol. 1 (pp. 7-8). Tallinn: Estonian Academy of Sciences.

Estonica, (2012, September 27). Eesti keel maailma taustal. Retrieved September 22, 2015 from http://www.estonica.org/et/%C3%9Chiskond/Eesti_keel/Eesti_keel_maailma_taustal/.

Face, T. L. & D'Imperio, M. (2005). Reconsidering a focal typology: Evidence from Spanish and Italian. *Rivista di Linguistica*, 17.2, 271-289.

Fanselow, B. & Lenertova, D. (2011). Left peripheral focus: mismatches between syntax and information structure. *Natural Language and Linguistic Theory*, 29(1), 160-209.

Féry, C. & Drenhaus, H. (2008). Single prosodic phrase sentences. In S. Ishihara, M. Schmitz & A. Schwarz (Eds.), *Interdisciplinary Studies in Information Structures 10. Working Papers of the SFB 632*(pp. 1-44). Potsdam.

Féry, C. & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics,* 36, 680-703.

Féry, C. & Krifka, M. (2008). Information structure. Notional distinctions, ways of expression. In P. Van Sterkenburg, (Ed.), *Unity and diversity of languages* (pp. 123-136). Amsterdam: John Benjamins.

Fillmore, Ch. J. (1985). Syntactic Intrusions and the Notion of Grammatical Construction. *Proceedings of the 11th Annual Meeting of the Berkeley Linguistics Society* (1985), (pp. 73-86).

Firbas, J. (1966). On defining the theme in functional sentence analysis. In Daneš et al. (Eds.) *Travaux Linguistiques de Prague* (Vol. 1, pp. 267-280). University of Alabama Press.

Fletcher, J. (2013). The Prosody of Speech: Timing and Rhythm. In W. Hardcastle, J. Laver & F. Gibbon (Eds.), *The Handbook of Phonetic Sciences, Second Edition* (pp. 523-602). Wiley-Blackwell Publishing.

Fowler, C.A., Levy, E.T. & Brown J.M. (1997) Reductions of Spoken Words in Certain Discourse Contexts. *Journal of memory and language*, 37, 24-40

Fowler, S. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech*, 31, 307-309

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of Acoustical Society of America* 27, 765–768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1, 126–152.

Gårding, E. (1981). Contrastive prosody: a model and its applications. *Studia Liguistica*, 35, 146-165.

Grice, M., Baumann, S., & Benzmüller, R., (2005). German intonation in autosegmental-metrical phonology. In Jun, S. (Ed.), *Prosodic Typology. The Phonology of Intonation and Phrasing* (pp. 55–83). Oxford: Oxford University Press.

Groenendijk, J. (1999). The logic of interrogation. In T. Matthews and D.L. Srolowitch (Eds.) *Semantics and Linguistic Theory* (pp. 109-126). Stanford: CLC Publications.

Gryllia, S. (2009). On the nature of preverbal Focus in Greek. A theoretical and experimental approach. Doctoral Thesis, University of Leiden. Utrecht: LOT

Gundel, J. (1985). 'Shared knowledge' and topicality. *Journal of Pragmatics*, 9, 83-107.

Gundel, J. (1999). On different kinds of focus. In Bosch, P. & Sandt, van der, R. (Eds.), *Focus. Linguistic, Cognitive, and Computational Perspectives* (pp. 293–305). Cambridge: Cambridge University Press.

Gussenhoven, C. (1983a). Focus, mode and the nucleus. *Journal of Linguistics*, 19, 377-417.

Gussenhoven, C. (1983b). Testing the reality of focus domains. *Language and speech*, 26, 61–80.

Gussenhoven, C. (1991). The English Rhythm Rule as an accent deletion rule. *Phonology*, 8, 1-35.

Gussenhoven, C. (1992). Sentence accents and argument structre. In I. M. Roca (Ed.), *Thematic Structure. Its Role in Grammar* (pp. 79-106). Berlin, New York: Foris Publications.

Gussenhoven, C. (1999). On the limits of focus projection in English. In Bosch, P. & Sandt, van der, R. (Eds.), *Focus. Linguistic, Cognitive, and Computational Perspectives* (pp. 43-54). Cambridge: Cambridge University Press.

Gussenhoven, C. (2007). Types of focus in English. In C. Lee, M. Gordon & D. Büring (Eds.), *Topic and Focus. Cross-Linguistic Perspectives on Meaning and Intonation. Studies in Linguistics and Philosophy*, 82, 83–100.

Halliday, M.A.K. (1967a). Notes on Transitivity and Theme in English: Part 2. *Journal of Linguistics*, 3(2), 199-244.

Halliday, M.A.K. (1967b). Intonation and grammar in British English. The Hague, Paris: Mouton.

Hanssen, J., Peters, J. & Gussenhoven, C., (2008). Prosodic effects of focus in Dutch declaratives. *Speech Prosody – Proceedings of the fourth Conference*, Campinas Brazil 2008 (pp. 609-612).

Haspelmath, M. (1997). Indefinite pronouns. Oxford: Oxford University Press.

Honda, K. (2004). Physiological factors causing tonal characteristics of speech: from global to local prosody. Speech Prosody, March 23-26, 2004, Nara, Japan.

Hovarth, J. (2010). "Discourse features", syntactic displacement and the status of contrast. *Lingua*, 120, 1346-1369.

Hualde, J. I. (2002). Intonation in Spanish and the other Ibero-Romance Languages. Overview and status questions. In Wiltshire, C. R. & J. Camps (Eds.), *Romance Phonology and Variation* (pp. 101-115). Amsterdam: John Benjamins B.V

Jelinek, E. (1984). Empty categories, case, and configurationality. *Natural Language and Linguistic Theory,* 2, 39-76.

Jokinen, K. (2005). A Grammar for Finnish Discourse Patterns. In Arppe et al. (Eds.) *Inquiries into Words, Constraints and Contexts. Festschrift for Kimmo Koskenniemi on his 60th Birthday* (pp. 227-240). CSLI Studies in Computational Linguistics Online.

Jun, S.-A. (2006). Prosodic Typology. In S.-A. Jun (Ed.), *Prosodic Typology. The Phonology of Intonation and Phrasing* (pp. 430-458). Oxford: Oxford University Press.

Kaiser & Trueswell (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, 94(2), 113-147.

Kasterpalu, R. (2013). Two different intonation contours of Estonian *jaajaa*. In E.-L. Asu & P. Lippus (Eds.), *Nordic Prosody – Proceedings of the XIth Conference*, Estonia, Tartu; 15.-17. Aug 2012 (pp. 177-186). Peter Lang Verlag.

Katz, J. & Selkirk, E. (2011). Contrastive focus *vs.* discourse-new: Evidence from phonetic prominence in English. *Language*, 87(4), 771-816.

Keevallik, L. (2003). Terminally Rising Pitch Contours of Response Tokens in Estonian. *Crossroads of Language, Interaction, and Culture*, Vol.5, 49-65.

Keller F. & Alexopoulou, T. (2001) Phonology competes with syntax: experimental evidence for the interaction of word order and accent placement in the realization of information structure. *Cognition*, 79(3), 301-372.

Knight, R.-A. (2003). Peaks and Plateaux: The production and perception of high intonational targets in English. Doctoral Dissertation, University of Cambridge

Kochanski G., Grabe, E., Coleman, J. and Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *Journal of Acoustical Society of America*, 118(2), 1038-1054.

Kohler, K. (1987a). The linguistic functions of F0 peaks. *Proceedings of the 11th International Congress of Phonetics Sciences*, Estonia, Tallinn 1-7 August 1987, (Vol. 3, 149-152).

Kohler, K. (1987b). Categorical pitch perception. *Proceedings of the 11th International Congress of Phonetics Sciences*, Estonia, Tallinn 1-7 August 1987, (Vol. 5, pp. 331-333).

Kohler, K. (1991). Form and function of intonation peaks in German; a research project. In K. J. Kohler, (Ed.), Studies in German Intonation (pg. 11-27). AIPUK 25.

Krahmer, E. & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication*, 34, 391-405.

Krifka, M. 2007. Basic notions of information structure. In C. Fery and M. Krifka (Eds.), Interdisciplinary Studies on Information Structure 6, (pp. 13-56). Potsdam: Universitätsverlag.

Kuno, S. (1972). Functional sentence perspective: a case study from Japanese and English. *Linguistic Inquiry*, 3, 269-320.

Laanekask, H. & Erelt, T. (2003). Written Estonian. In M. Erelt (Ed.), *The Estonian language. Linguistica Uralica. Supplementary Series I*, (pp. 273–342). Tallinn: Eesti Teaduste Akadeemia Kirjastus.

Ladd, D. R. (1986). Intonational Phrasing: The Case of Recursive Prosodic Structure. *Phonology Yearbook*, 3, 311-340.

Ladd, D. R. (2001). Intonation. In M. Haspelmath, E. König, W. Oesterreicher, W. Raible (Eds.), *Language typology and language universals: an international handbook*, (Vol. 2, pp. 1380-1390). Berlin: Walter de Gruyter.

Ladd, D. R. (2008). Intonational phonology. Cambridge Studies in Linguistics, 119. Cambridge: Cambridge University Press.

Ladd, D. R. & Morton, R. (1997). The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics*, 25, 313-342.

Ladd, D. R., Faulkner, D., Faulkner, H. & Schepman, A. (1999). Constant 'segmental' anchoring of f0 movements under changes in speech rate. *Journal of the Acoustical Society of America*, 106, 1543-1554.

Ladd, D.R., Mennen, I., & 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: Cambridge University Press.

Lehiste, I. (1960). Segmental and syllabic quantity in Estonian. In T. A. Sebeok (Ed.), *Uralic and Altaic Series. American Studies in Uralic Linguistics*, 1, 21–82.

Lehiste, I. (1965). The Function of Quantity in Finnish and Estonian. *Language*, 41 (3), 447-456.

Lehiste, I. (1970). Suprasegmentals. Cambridge, Massachusetts: The Massachusetts Institute of Technology Press

Lehiste, Ilse (1997). Search for Phonetic Correlates in Estonian Prosody. In I. Lehiste & J. Ross (Eds.), *Estonian Prosody: Papers from a Symposium* (pp. 11–35).

Lehiste, I. (2002). Prosodic change in progress: from quantity language to accent language. In P. Fikkert & H. Jacobs (Eds.), *Development in Prosodic Systems* (pp. 47-65). Berlin, New York: Mouton de Gruyter.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of Acoustical Society of America*, 32, 451-454.

Lindström, L. (2002). Veel kord subjekti ja predikaadi vastastikusest asendist laiendi järel. *Emakeele Seltsi aastaraamat* (pp. 87-105). Tartu: Emakeele Selts.

Lindström, L. (2004). Sõnajärg lause tuumaargumentide eristajana eesti keeles. In L. Lindström (Ed.), *Lauseliikmeist eesti keeles. Tartu ülikooli eesti keele õppetooli preprindid* (pp. 40–49). Tartu: Tartu Ülikool.

Lindström, L. (2006). Infostruktuuri osast eesti keele sõnajärje muutumisel. (On the role of the information structure in the change of the Estonian word order). *Keel ja Kirjandus*, 11, 875–888. (Language and Literature, 11, 875–888).

Lippus, P., Pajusalu, K. & Allik, J. (2009). The tonal component of Estonian quantity in native and non-native perception. *Journal of Phonetics*, 37, 388–396.

Lippus, P., Pajusalu, K. & Allik, J. (2011). The role of pitch cue in the perception of the Estonian long quantity. In S. Frota, G. Elordieta & P. Prieto (Eds.), *Prosodic Categories: Production, Perception and Comprehension* (pp. 231-242). Springer.

Lippus, P., Asu, E. L., Teras, P., Tuisk, T. (2013). Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian. *Journal of Phonetics*, 41(1), 17 - 28.

Lippus, P. Asu, E. L. & Kalvik, M.-L. (2014). An acoustic study of Estonian word stress. Campbell, N., Gibbon, D. & Hirst, D. (Eds.). *Social and Linguistic Speech Prosody. Proceedings of the 7th international conference on Speech Prosody* (pp. 232 - 235). Dublin: International Speech Communications Association.

Maeda, S. (1976). A Characterization of American English Intonation. Doctoral dissertation. Massachusetts Institute of Technology.

Merin, A. (1994). Algebra on elementary social acts. In S. L. Tsohatzidis (Ed.), Foundations of speech act theory. Philosophical and linguistic perspectives (pp. 234–266). London: Routledge.

Mihkla, M. & Kalvik M.-L., (2011). Significant features in Estonian word prosody. *Proceedings of XVIIth International Conference of Phonetic Sciences*, Hong Kong 17-21 August.

Mithun, M. (1987). Is basic word order universal? In R. S. Tomlin (Ed.), *Coherence and grounding in discourse. Outcome of a Symposium*, Eugene, Oregon, June 1984. Amsterdam/Philadelphia: John Benjamins Publishing Company

Mycock, L. (2010). Prominence in Hungarian: the prosody-syntax connection. *Transactions of the Philological Society*, 108(3), 265-297.

Nespor, M. & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris. Republished 2007 by Mouton de Gruyter, Berlin.

Niebuhr, O. (2007). Perzeption und kognitive Verarbeitung der Sprechmelodie. Theoretische Grundlagen und empirische Untersuchungen (Language, Context, and Cognition, Vol. VII). Berlin/New York: deGruyter.

Pajusalu, K. (2003). Estonian dialects. In M. Erelt (Ed.), The Estonian language. *Linguistica Uralica. Supplementary Series I*, (pp. 231–272). Tallinn: Eesti Teaduste Akadeemia Kirjastus.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of Acoustical Society of America*, 66(2), 363-369.

Pierrehumbert, Janet B. (1980). The phonology and phonetics of English intonation. Doctoral dissertation, The Massachusetts Institute of Technology.

Pierrehumbert, J. & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan & M.E. Pollack (Eds.), *Intensions in communication* (pp. 271-311). Cambridge, Massachusetts: The MIT Press.

Pierrehumbert, J. B. & Steele, S.A. (1989). Categories of tonal alignment in English. *Phonetica*, 46, 181-196.

Plüschke, M. (2011). Peak alignment in falling accents in Estonian. *Proceedings of XVIIth International Conference of Phonetic Sciences*, Hong Kong 17-21 August.

Plüschke, M. (2013). Peak Alignment in Estonian. Doctoral dissertation, Ludwig-Maximilians-Universität München.

Prieto, P., van Santen, J., and Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, 23, 429–451.

Primus, B. (2001). Word order typology. In M. Haspelmath, E. König, W. Oesterreicher, W. Raible (Eds.), *Language typology and language universals: an international handbook*, (Vol. 2, pp. 855-873). Berlin: Walter de Gruyter.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P.Cole, (Ed.), *Radical Pragmatics*, (pp. 223-255). New York: Academic Press

Prince, E.G. (1992). The ZPG letter: Subjects, definiteness, and information-status. In Thompson, S. & Mann, W. (Eds.), *Discourse Description: Diverse Analyses of a Fund Raising Text*, (pp. 295-325). Philadelphia: John Benjamins.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rathcke, T. & Harrington, J. (2006). Is there a distinction between H+!H* and H+L* in standard German? Evidence from an acoustic and auditory analysis. *Speech Prosody – Proceedings of the third International Conference*, Dresden, 2-5 May 2006. Dresden: TUDPress.

Reinhart, T. (1981). Pragmatics and linguistics: an analysis of sentence topics. *Philosophica*, 27(1), 53-94.

Remmel, N. (1963). Sõnajärjestus eesti lauses. Deskriptiivne käsitlus. *Eesti keele süntaksi küsimusi* (pp. 216-217). Tallinn: Eesti Riiklik Kirjastus.

Repp, S. (2010). Defining contrast as an information-structural notion in grammar. Editorial. *Lingua*, 120, 1333-1345.

Rietveld, A.C.M & Gussenhoven, C. (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics*, 13, 299-308.

Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of Grammar. Handbook in generative syntax* (pp. 281-337). Dordrecht: Kluwer Academic Publishers.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1, 75–116.

Sahkai, H., Kalvik, M.-L., Mihkla, M. (2013). Prosody of contrastive focus in Estonian. *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, Lyon, France, August 25-29, 2013 (315-319). International Speech Communication Association.

Sahkai, H., Kalvik, M.-L., Mihkla, M. (2013b). Prosodic effects of the information structure in Estonian. In E. L. Asu & P. Lippus (Eds.), *Nordic Prosody – Proceedings of the XIth Conference*, Tartu 2012 (pp. 323-332). Peter Lang Verlag.

Salveste, N. (2010). Variation of pitch cues in the perception of Estonian quantities. In: Werner, S. & Kinnunen, T. (Ed.), Proceedings of the XXVI Phonetics *Fonetiikan Päivät 2010 – The Phonetics Symposium*, Finland, Mekrijärve, 2010. University of Eastern Finland.

Salveste, N. (2013). Focus perception in Estonian: Is it governed by syntax or by prosody? In E. L. Asu & P. Lippus (Eds.), *Nordic Prosody – Proceedings of the XIth Conference*, Tartu 2012 (pp. 333-342). Peter Lang Verlag.

Salveste, N., Harrington, J. & Kleber, F. (2015). Phonetic effects of corrective focus in Estonian. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, August 6-10, 2015.

Samek-Lodovici, V. (2005). Prosody-syntax interaction in the expression of focus. *Natural Language & Linguistic Theory*, 23, 687-755.

Schwarzchild, R. (1999). Givenness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics*, 7(2), 141–177.

Selkirk, E. (1984), Phonology and Syntax: The Relation between Sound and Structure. Cambridge, Massachusetts: MIT Press.

Selkirk, E. (1995), Sentence Prosody: Intonation, Stress, and Phrasing, In J. A. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 550-569). London: Basil Blackwell.

Sgall, P., Hajicová, E. & Panevová, J. (1986). The Meaning of the Sentence in Its Semantic and Pragmatic Aspects. Prague: D. Reidel Publishing Company.

Siewierska, A. (1998). Variation in major constituent order: a global and a European perspective. In A. Siewierska (Ed.), *Constituent Order in the Languages of Europe* (pp. 475-551). Berlin: Mouton de Gruyter

Silverman, K. & J. Pierrehumbert (1990). The Timing of Prenuclear High Accents in English. *Papers in Laboratory Phonology I* (pp. 72-106). Cambridge: Cambridge University Press.

Simpson, J. (1983). Aspects of Warlpiri morphology and syntax. Doctoral dissertation, The Massachusetts Institute of Technology.

Siptár, P. & Terköczy, M. (2000). The phonology of Hungarian. Oxford: Oxford University Press.

Skopeteas, S. & Fanselow, G. (2009). Effects of givenness and constraints on free word order. In M. Zimmermann & C. Féry (Eds.), *Information structure* (pp. 307-331). Oxford: Oxford University Press.

Skopeteas S., Féry, C. and Asatiani, R. (2009). Word order and intonation in Georgian. *Lingua*, 119, 102-127.

Sluijter, A. & van Heuven, V. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of Acoustic Society of America*, 100, 2471-2485.

Steele, S. (1978). Word order variation: A typological study. In Ch. Greenberg (Ed), *Universals of human language. Syntax*, 4, 585-624.

Stechow, A. (1989). Focusing and backgrounding operators. Technical report, Universität Konstanz.

Suomi, K., Toivanen, J. & Ylitalo, R. (2003). Durational and tonal correlates of accent in Finnish. *Journal of Phonetics*, 31, 113-138.

Swerts, M., Strangert, E. and Heldner, M. (1996). F0 declination in read-aloud and spontaneous speech. *Proceedings of the fourth International Conference on Spoken Language Processing*, USA, Philadelphia, October 3-6, 1996.

Swerts, M., Krahmer, E. & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: a comparative analysis. *Journal of Phonetics*, 30, 629-654.

Szendröi, K. (2003). A stress-based approach to the syntax of Hungarian focus. *The Linguistic Review*, 20, 37-78.

t' Hart, J. Collier, R. & Cohen, A. (1990). A perceptual study of intonation. An experimental-phonetic approach to speech melody. Cambridge: Cambridge University Press.

Tael, K. (1988) Sõnajärjemallid eesti keeles (võrrelduna soome keelega). *Preprint KKI-56*. Tallinn: Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.

Terken, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America*, 89, 1768-76.

Terken, J. & Hirschberg, J. (1994). Deaccentuation of words representing 'given' information: effects of persistence of grammatical role and surface position. *Language and Speech*, 37 (2), 125–145.

Titov, E. (2012). Information structure of argument order alternations. Doctoral dissertation, University College London.

Truckenbrodt, H. & Darcy, I. (2010). Object clauses, movement and phrasal stress. In N. Erteschik-Shir & L. Rochman (Eds.), *The sound patterns of syntax* (pp. 189-216). Oxford: Oxford University Press.

Turk, A. E. & Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *Journal of Acoustical Society of America*, 99, 3782-3790.

Turk, A. E. & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, 27(2), 171-206.

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. *Journal of the Acoustical Society of America*, 121(2), EL 55–61.

Välimaa-Blum, R. M. (1988). Finnish existential clauses - their syntax, pragmatics and intonation. Doctoral dissertation, The Ohio State University.

Välimaa-Blum, R. M. (1993). Intonation: a distinctive parameter in grammatical constructions. *Phonetica*, 50, 124-137.

Vallduví, E. (1993). The Informational Component. IRCS Technical Reports Series. University of Pennsylvania, Institute for Research in Cognitive Science. <http://repository.upenn.edu/ircs reports/188>

Vallduví, E. & Engdahl, E. (1996). The linguistic realisation of information packaging. *Linguistics*, 34(3), 459–520.

Van Valin, R. D. (1993). A synopsis of Role and Reference Grammar. In R. D. Van Valin (Ed.), *Advances in Role and Reference Grammar* (pg. 1-164). Amsterdam: Benjamins.

Van Valin, R. D. (1999). A typology of the interaction of focus structure and syntax. In Raxilina, E. & Testelec, J. (Eds.), *Typology and the Theory of Language: From Description to Explanation* (pp. 511-524). Moscow: Languages of Russian Culture.

Van Valin Jr., R. D. (2009). Case in role and reference grammar. In A. Malchukov, & A. Spencer (Eds.), The Oxford handbook of case (pp. 102-120). Oxford University Press.

Varga, L. (2002). Intonation and Stress. Evidence from Hungarian. New York: Palgrave Macmillian.

Vende, K. (1975). Intonation of question and answer in Estonian: I. In A. Eek (Ed.), *Estonian Papers on Phonetics* (pp. 86-113). Tallinn: Academy of Sciences of the Estonian S.S.R.

Vende, (1982). Intonation of question and answer in Estonian: II. Synthetic stimuli. In A. Eek (Ed.), *Estonian Papers on Phonetics* (pp. 77-106). Tallinn: Academy of Sciences of the Estonian S.S.R.

Viitso, T.-R. (2003). Structure of the Estonian language. Phonology, morphology and word formation. *Estonian language. Linguistica Uralica, Supplementary series* (Vol. 1, pp. 9-92). Tallinn: Estonian Academy of Sciences.

Vilkuna, M. (1995). Discourse configurationality in Finnish. In K. É. Kiss (Ed.), *Discourse configurational languages* (pp. 244-268). New York: Oxford University Press.

Vilkuna, M. (1998). Word order in European Uralic. In A. Siewierska (Ed.), *Constituent Order in the Languages of Europe* (pp. 173-233). Berlin: Mouton de Gruyter.

Ward, G. and Birner, B. J. (2004). Information Structure. In L. R. Horn & G. Ward (Eds.), *Handbook of Pragmatics* (pp. 153-174). Oxford: Basil Blackwell.

Weskott, T., Hörnig, R., Fanselow, G. & Kliegl, R. (2011). Contextual licensing of marked OVS word order in German. *Linguistische Berichte*, 225, 3-18.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33(2), 159-197.

Zimmermann, M. & Onea, E. (2011). Focus marking and focus interpretation. Lingua, 121, 1651-1670.

Zubizarreta, M. L. (1998). Prosody, Focus, and Word Order. Cambridge, Massachusetts: MIT-Press.

Yuan, J. & Liberman, M. (2014). F0 declination in English and Mandarin Broadcast News Speech. *Speech Communication*, 65, 67-74.

Appendix A: Experimental materials

A.1. Focus perception in Estonian: syntactic or prosodic

A context was written on the upper part of the computer screen. See Figure 2.2 for an experimental assignment. Participants were asked to read the context and after this, listen to the two stimuli (stimuluspairs: a|b, a|c, a|d, b|c, b|d, c|d) and choose the stimulus that was most appropriate to the context.

Focus in Context	**Context**	**Target**	**Stimulus**	**Type**	**Word order**	**Placement of Nuclear Accent**
Object NP	Verandal seisavad värvid. Õues on ilus ilm ja mõtlesime, mida me õuel värvida võiksime.	Laeva	Värvima õuel LAEVA	a.	VAO	Final
			Värvime ÕUEL laeva	b.	VOA	Medial
Adverb	Verandal seisavad värvid. Vanaisa tõi välja vana roostes laeva ja mõtlesime, kus me laeva värvida võiksime.		Värvime LAEVA õuel	c.	VAO	Medial
			Värvime laeva ÕUEL	d.	VOA	Final
Object NP	Verandal seisid värvid. Õues oli ilus ilm ja mõtlesime, mida me värvida võiksime.	Raami	Värvime õuel RAAMI	a.	VAO	Final
			Värvime RAAMI õuel	b.	VOA	Medial
Adverb	Verandal seisid värvid. Vanaisa tõi koju vana puitraami ja mõtlesime, kus me raami värvida võiksime.		Värvime ÕUEL raami	c.	VAO	Medial
			Värvime raami ÕUEL	d.	VOA	Final
Object NP	Aidas seisid maalipintslid. Rannal oli tuulevaikne ja mõtlesime, mida me seal maalida võiksime.	Muuli	Maalime rannal muuli	a.	VAO	Final
				b.	VOA	Medial
Adverb	Aidas seisid maalipintslid. Saarel oli palju karjamaid ja mõtlesime, kus me loomi maalida võiksime.			c.	VAO	Medial
				d.	VOA	Final
Object NP Adverb	Aidas seisid maalipintslid. Rannal oli tuulevaikne ja mõtlesime, mida me seal maalida võiksime. Aidas seisid maalipintslid. Saarel oli palju huvitavaid lilli ja mõtlesime, kus me lilli maalida võiksime.	Lilli	Maalime rannal lilli	a. b. c.	VAO VOA VAO	Final Medial Medial

				d.	VOA	Final
Object NP	Keldris seisis savi. Päike oli muuli soojaks kütnud ja mõtlesime, mida me muulil voolida võiksime.	Loomi	Voolime muulil loomi	a.	VAO	Final
				b.	VOA	Medial
Adverb	Keldris seisis savi. Nooremale õele meeldisid hirmsasti savist loomad ja mõtlesime, kus me loomi voolida võiksime.			c.	VAO	Medial
				d.	VOA	Final
Object NP	Keldris seisis savi. Päike oli muuli soojaks kütnud ja mõtlesime, mida me muulil voolida võiksime.	Vaala	Voolime muulil vaala	a.	VAO	Final
				b.	VOA	Medial
Adverb	Keldris seisis savi. Nooremale õele meeldisid hirmsasti savist vaalad ja mõtlesime, kus me loomi voolida võiksime.			c.	VAO	Medial
				d.	VOA	Final
Object NP	Aidas seisavad liimid.Õuel seisis palju parandamist vajavaid asju ja mõtlesime, mida me liimida võiksime.	Tooli	Liimime õuel tooli	a.	VAO	Final
				b.	VOA	Medial
Adverb	Aidas seisid liimid.Vanaisa tõi koju katkist mööblit ja mõtlesime, kus me tooli liimida võiksime.			c.	VAO	Medial
				d.	VOA	Final
Object NP	Aidas seisavad liimid.Õuel seisis palju parandamist vajavaid asju ja mõtlesime, mida me liimida võiksime.	Maali	Liimime õuel maali	a.	VAO	Final
				b.	VOA	Medial
Adverb	Aidas seisid liimid.Vanaisa tõi koju vanu kunstitöid ja mõtlesime, kus me maali liimida võiksime.			c.	VAO	Medial
				d.	VOA	Final
Object NP	Kuuris seisavad loodid.Väike Vello uuris, mida loodiga teha ja mõtlesime, mida me õuel loodida võiksime.	Müüri	Loodime rannal müüri	a.	VAO	Final
				b.	VOA	Medial
Adverb	Kuuris seisavad loodid.Väike Vello uuris, kuidas müüri tasapinda hinnata ja mõtlesime, kus me müüri loodida võiksime.			c.	VAO	Medial
				d.	VOA	Final
Object NP	Kuuris seisavad loodid.Väike Vello uuris,	Lauda	Loodime õuel lauda	a.	VAO	Final

	mida loodiga teha ja mõtlesime, mida me rannal loodida võiksime.			b.	VOA	Medial
Adverb	Kuuris seisavad loodid.Väike Vello uuris, kuidas laevakabiini tasapinda hinnata ja mõtlesime, kus me laeva loodida võiksime.			c. d.	VAO	Medial Final

A.2. Broad and narrow focus in Estonian

The participants were asked to utter target-sentences as an appropriate response to the context that was played over loudspeakers and written on the upper part of the computer screen. The target sentences were varied with the filler sentences.

Function	Focus Type	Focus Position	Word order	Word in Focus	Context	Sentence
Target	Broad		SVO		Mis uudist?	Leena maalis vaala.
						Miili kuulis Eevat.
						Meeli hüüdis Loonat.
						Liina liimis raami.
			OVS			Vaala maalis Leena.
						Eevat kuulis Miili.
						Loonat hüüdis Meeli.
						Raami liimis Liina.
	Narrow	Initial	SVO	Leena	Keegi ju maalis vaala?	Leena maalis vaala.
				Miili	Keegi ju kuulis Eevat?	Miili kuulis Eevat!
				Meeli	Keegi ju hüüdis Loonat?	Meeli hüüdis Loonat!
				Liina	Keegi ju liimis raami?	Liina liimis raami!
			OVS	vaala	Leena ju maalis midagi?	Vaala maalis Leena.
				Eevat	Miili ju kuulis kedagi?	Eevat kuulis Miili.
				Loonat	Meeli ju hüüdis kedagi?	Loonat hüüdis Meeli.
				raami	Liina ju liimis midagi?	Raami liimis Liina.
		Final	SVO	vaala	Leena ju maalis midagi?	Leena maalis vaala.
				Eevat	Miili ju kuulis kedagi?	Miili kuulis Eevat!
				Loonat	Meeli ju hüüdis kedagi?	Meeli hüüdis Loonat!
				raami	Liina ju liimis midagi?	Liina liimis raami!
			OVS	Leena	Keegi ju maalis vaala?	Vaala maalis Leena.
				Miili	Keegi ju kuulis Eevat?	Eevat kuulis Miili.
				Meeli	Keegi ju hüüdis Loonat?	Loonat hüüdis Meeli.
				Liina	Keegi ju liimis raami?	Raami liimis Liina.
Filler	Broad		AVO		Mis uudist?	Laadal müüdi moone.
						Rannas joodi veini.

						Õuel nähti loomi.
						Kuuris löödi raudu.
			OVA			Moone müüdi laadal.
						Veini joodi rannas.
						Loomi nähti õuel.
						Raudu löödi kuuris.
	Narrow	Initial	AVO	laadal	Kuskil ju müüdi moone?	Laadal müüdi moone.
				rannas	Kuskil ju joodi veini?	Rannas joodi veini.
				õuel	Kuskil ju nähti loomi?	Õuel nähti loomi.
				kuuris	Kuskil ju löödi raudu?	Kuuris löödi raudu.
			OVA	moone	Laadal müüdi ju midagi?	Moone müüdi laadal.
				veini	Rannas ju joodi midagi?	Veini joodi rannas.
				loomi	Õuel ju nähti kedagi?	Loomi nähti õuel.
				raudu	Midagi ju löödi kuuris?	Raudu löödi kuuris.
		Final	AVO	moone	Laadal müüdi ju midagi?	Laadal müüdi moone.
				veini	Rannas ju joodi midagi?	Rannas joodi veini.
				loomi	Õuel ju nähti kedagi?	Õuel nähti loomi.
				raudu	Midagi ju löödi kuuris?	Kuuris löödi raudu.
			OVA	laadal	Kuskil ju müüdi moone?	Moone müüdi laadal.
				rannas	Kuskil ju joodi veini?	Veini joodi rannas.
				õuel	Kuskil ju nähti loomi?	Loomi nähti õuel.
				kuuris	Kuskil ju löödi raudu?	Raudu löödi kuuris.

A.3. Corrective focus in Estonian

The participants were asked to utter target-sentences as an appropriate response to the context that was played over loudspeakers and written on the upper part of the computer screen. The target sentences were varied with the filler sentences.

Function	Focus Type	Focus Position	Word order	Word in Focus	Context	Sentence
Target	New	Initial	SVO	Leena	Keegi ju maalis vaala?	Leena maalis vaala.
				Miili	Keegi ju kuulis Eevat?	Miili kuulis Eevat.
				Meeli	Keegi ju hüüdis Loonat?	Meeli hüüdis Loonat.
				Liina	Keegi ju liimis raami?	Liina liimis raami.
			OVS	vaala	Leena ju maalis midagi?	Vaala maalis Leena.
				Eevat	Miili ju kuulis kedagi?	Eevat kuulis Miili.
				Loonat	Meeli ju hüüdis kedagi?	Loonat hüüdis Meeli.
				raami	Liina ju liimis midagi?	Raami liimis Liina.
		Final	SVO	vaala	Leena ju maalis midagi?	Leena maalis vaala.
				Eevat	Miili ju kuulis kedagi?	Miili kuulis Eevat.
				Loonat	Meeli ju hüüdis kedagi?	Meeli hüüdis Loonat.
				raami	Liina ju liimis midagi?	Liina liimis raami.
			OVS	Leena	Keegi ju maalis vaala?	Vaala maalis Leena.
				Miili	Keegi ju kuulis Eevat?	Eevat kuulis Miili.
				Meeli	Keegi ju hüüdis Loonat?	Loonat hüüdis Meeli.
				Liina	Keegi ju liimis raami?	Raami liimis Liina.
	Correc-tive	Initial	SVO	Miili	Aivo kuulis Eevat.	Miili kuulis Eevat.
				Meeli	Reena hüüdis Loonat.	Meeli hüüdis Loonat.
				Liina	Taavi liimis raami.	Liina liimis raami.
				Leena	Anna maalis vaala.	Leena maalis vaala.
			OVS	Eevat	Miili kuulis Meerit.	Eevat kuulis Miili.
				Loonat	Meeli hüüdis Taavit.	Loonat hüüdis Meeli.

				Raami	Liina liimis kannu.	Raami liimis Liina.
				Vaala	Leena maalis moone.	Vaala maalis Leena.
		Final	SVO	Eevat	Miili kuulis Meerit.	Miili kuulis Eevat.
				Loonat	Meeli hüüdis Taavit.	Meeli hüüdis Loonat.
				raami	Liina liimis kannu.	Liina liimis raami.
				vaala	Leena maalis moone.	Leena maalis vaala.
			OVS	Miili	Aivo kuulis Eevat.	Eevat kuulis Miili.
				Meeli	Reena hüüdis Loonat.	Loonat hüüdis Meeli.
				Liina	Taavi liimis raami.	Raami liimis Liina.
				Leena	Anna maalis vaala.	Vaala maalis Leena.
Filler	New	Initial	AVO	laadal	Kuskil ju müüdi moone?	Laadal müüdi moone.
				rannas	Kuskil ju joodi veini?	Rannas joodi veini.
				õuel	Kuskil ju nähti loomi?	Õuel nähti loomi.
				kuuris	Kuskil ju löödi raudu?	Kuuris löödi raudu.
			OVA	moone	Laadal müüdi ju midagi?	Moone müüdi laadal.
				veini	Rannas ju joodi midagi?	Veini joodi rannas.
				loomi	Õuel ju nähti kedagi?	Loomi nähti õuel.
				raudu	Midagi ju löödi kuuris?	Raudu löödi kuuris.
		Final	AVO	moone	Laadal müüdi ju midagi?	Laadal müüdi moone.
				veini	Rannas ju joodi midagi?	Rannas joodi veini.
				loomi	Õuel ju nähti kedagi?	Õuel nähti loomi.
				raudu	Midagi ju löödi kuuris?	Kuuris löödi raudu.
			OVA	laadal	Kuskil ju müüdi moone?	Moone müüdi laadal.
				rannas	Kuskil ju joodi veini?	Veini joodi rannas.
				õuel	Kuskil ju nähti loomi?	Loomi nähti õuel.
				kuuris	Kuskil ju löödi raudu?	Raudu löödi kuuris.
	Correc-tive	Initial	AVO	Laadal	Rannas müüdi moone.	Laadal müüdi moone.
				Rannas	Lehtlas joodi veini.	Rannas joodi veini.
				Õuel	Rannas nähti loomi.	Õuel nähti loomi.
				Kuuris	Aidas löödi raudu.	Kuuris löödi raudu.
			OVA	Moone	Rannas müüdi moone.	Moone müüdi laadal.

				Veini	Lehtlas joodi veini.	Veini joodi rannas.
				Loomi	Rannas nähti loomi.	Loomi nähti õuel.
				Raudu	Aidas löödi raudu.	Raudu löödi kuuris.
		Final	AVO	moone	Laadal müüdi tulpe.	Laadal müüdi moone.
				veini	Rannas joodi mahla.	Rannas joodi veini.
				loomi	Õuel nähti rahvast.	Õuel nähti loomi.
				raudu	Kuuris löödi kirveid.	Kuuris löödi raudu.
			OVA	laadal	Laadal müüdi tulpe.	Moone müüdi laadal.
				rannas	Rannas joodi mahla.	Veini joodi rannas.
				õuel	Õuel nähti rahvast.	Loomi nähti õuel.
				kuuris	Kuuris löödi kirveid.	Raudu löödi kuuris.

A.4. Phonetics of givenness in Estonian

Participants were asked to read out aloud target sentences that were presented in blocks of sentences. The blocks of target sentences were varied with the blocks of filler sentences.

Func.	Group	Nr. of Block	Target	Grammar of Target	Word order	Inf. str.	Target sentence
Targets	1	1	diiva	Object NP	SVO	New	Ööbik kuulab diivat.
						Given	Dirigent kuulab diivat.
						Given	Bariton kuulab diivat.
						Given	Taavi kuulab diivat.
		2	leedi			New	Härra kuulab leedit.
						Given	Poiss kuulab leedit.
						Given	Assistent kuulab leedit.
						Given	Riina kuulab leedit.
		3	muusa			New	Proua kuulab muusat.
						Given	Munk kuulab muusat.
						Given	Preester kuulab muusat.
						Given	Viive kuulab muusat.
		4	piiga			New	Emme kuulab piigat.
						Given	Müüja kuulab piigat.
						Given	Kelner kuulab piigat.
						Given	Loona kuulab piigat.
		5	beebi	Subject NP	OVS	New	Emmet kuulab beebi.
						Given	Puumat kuulab beebi.
						Given	Issit kuulab beebi.
						Given	Liinat kuulab beebi.
		6	joogi			New	Härrat kuulab joogi.
						Given	Publikut kuulab joogi.
						Given	Kardinali kuulab joogi.
						Given	Tiinat kuulab joogi.
		7	laama			New	Prouat kuulab laama.
						Given	Ministrit kuulab laama.
						Given	Saadikut kuulab laama.
						Given	Leenat kuulab laama.
		8	liige			New	Hindut kuulab liige.
						Given	Kohtunikku kuulab liige.
						Given	Juhatajat kuulab liige.
						Given	Jaanat kuulab liige.
		9	beebi	Object NP	OVS	New	Beebit kuulab puuma.
						Given	Beebit kuulab vanaisa.

						Given	Beebit kuulab issi.
						Given	Beebit kuulab Liina.
		10	joogi			New	Joogit kuulab neiu.
						Given	Joogit kuulab publik.
						Given	Joogit kuulab kardinal.
						Given	Joogit kuulab Tiina.
		11	laama			New	Laamat kuulab mamma.
						Given	Laamat kuulab minister.
						Given	Laamat kuulab piloot.
						Given	Laamat kuulab Leena.
		12	liige			New	Liiget kuulab kunde.
						Given	Liiget kuulab kohtunik.
						Given	Liiget kuulab misjonär.
						Given	Liiget kuulab Jaana.
		13	diiva	Subject NP	SVO	New	Diiva kuulab härrat.
						Given	Diiva kuulab dirigenti.
						Given	Diiva kuulab baritoni.
						Given	Diiva kuulab Taavit.
		14	leedi			New	Leedi kuulab kaasat.
						Given	Leedi kuulab poega.
						Given	Leedi kuulab assistenti.
						Given	Leedi kuulab Riinat.
		15	muusa			New	Muusa kuulab neiut.
						Given	Muusa kuulab munka.
						Given	Muusa kuulab preestrit.
						Given	Muusa kuulab Viivet.
		16	piiga			New	Piiga kuulab puumat.
						Given	Piiga kuulab lauljat.
						Given	Piiga kuulab kelnerit.
						Given	Piiga kuulab Loonat.
	2	17	beebi	Object NP	SVO	New	Emme kuulab beebit.
						Given	Vanaema kuulab beebit.
						Given	Tiiger kuulab beebit.
						Given	Liina kuulab beebit.
		18	joogi			New	Proua kuulab joogit.
						Given	Filosoof kuulab joogit.
						Given	Prostituut kuulab joogit.
						Given	Tiina kuulab joogit.
		19	laama			New	Hindu kuulab laamat.
						Given	President kuulab laamat.
						Given	Külaline kuulab laamat.
						Given	Leena kuulab laamat.
		20	liige			New	Naaber kuulab liiget.
						Given	Direktor kuulab liiget.
						Given	Pastor kuulab liiget.

					Given	Jaana kuulab liiget.
	21	diiva	Subject NP	OVS	New	Kaasat kuulab diiva.
					Given	Kirjanikku kuulab diiva.
					Given	Portjeed kuulab diiva.
					Given	Taavit kuulab diiva.
	22	leedi			New	Härrat kuulab leedi.
					Given	Autojuhti kuulab leedi.
					Given	Naabrit kuulab leedi.
					Given	Riinat kuulab leedi.
	23	muusa			New	Kääbust kuulab muusa.
					Given	Vanameest kuulab muusa.
					Given	Bussijuhti kuulab muusa.
					Given	Viivet kuulab muusa.
	24	piiga			New	Kerjust kuulab piiga.
					Given	Diktorit kuulab piiga.
					Given	Õpetajat kuulab piiga.
					Given	Loonat kuulab piiga.
	25	diiva	Object NP	OVS	New	Diivat kuulab muusik.
					Given	Diivat kuulab kirjanik.
					Given	Diivat kuulab portjee.
					Given	Diivat kuulab Taavi.
	26	leedi			New	Leedit kuulab neiu.
					Given	Leedit kuulab autojuht.
					Given	Leedit kuulab reporter.
					Given	Leedit kuulab Riina.
	27	muusa			New	Muusat kuulab härra.
					Given	Muusat kuulab vanamees.
					Given	Muusat kuulab bussijuht.
					Given	Muusat kuulab Viive.
	28	piiga			New	Piigat kuulab emme.
					Given	Piigat kuulab diktor.
					Given	Piigat kuulab õpetaja.
					Given	Piigat kuulab Loona.
	29	beebi	Subject NP	SVO	New	Beebi kuulab puumat.
					Given	Beebi kuulab vanaema.
					Given	Beebi kuulab emmet.
					Given	Beebi kuulab Liinat.
	30	joogi			New	Joogi kuulab hindut.
					Given	Joogi kuulab prostituuti.
					Given	Joogi kuulab jüngrit.
					Given	Joogi kuulab Tiinat.
	31	laama			New	Laama kuulab prouat.
					Given	Laama kuulab presidenti.
					Given	Laama kuulab külalist.
					Given	Laama kuulab Leenat.

		32	liige			New	Liige kuulab kundet.
						Given	Liige kuulab direktorit.
						Given	Liige kuulab pastorit.
						Given	Liige kuulab Jaanat.
Fillers	1	1	giidid	Object NP	AVO	New	Turule viiakse giidid.
						Given	Laadale viiakse giidid.
						Given	Hotelli viiakse giidid.
						Given	Randa viiakse giidid.
		2	härrad			New	Turule viiakse härrad.
						Given	Laadale viiakse härrad.
						Given	Kirikusse viiakse härrad.
						Given	Randa viiakse härrad.
		3	siilid			New	Turule viiakse siilid.
						Given	Laadale viiakse siilid.
						Given	Metsa viiakse siilid.
						Given	Randa viiakse siilid.
		4	vaalad			New	Turule viiakse vaalad.
						Given	Laadale viiakse vaalad.
						Given	Näitusele viiakse vaalad.
						Given	Randa viiakse vaalad.
		5	koorid	Subject NP	AVS	New	Turul vahivad koorid.
						Given	Laadal vahivad koorid.
						Given	Saalis vahivad koorid.
						Given	Rannas vahivad koorid.
		6	loomad			New	Turul vahivad loomad.
						Given	Laadal vahivad loomad.
						Given	Laudas vahivad loomad.
						Given	Rannas vahivad loomad.
		7	noored			New	Turul vahivad noored.
						Given	Laadal vahivad noored.
						Given	Teatris vahivad noored.
						Given	Rannas vahivad noored.
		8	tiimid			New	Turul vahivad tiimid.
						Given	Laadal vahivad tiimid.
						Given	Konverentsil vahivad tiimid.
						Given	Rannas vahivad tiimid.
		9	koorid	Object NP	OVA	New	Koorid viiakse turule.
						Given	Koorid viiakse laadale.
						Given	Koorid viiakse saali.
						Given	Koorid viiakse randa.
		10	loomad			New	Loomad viiakse turule.
						Given	Loomad viiakse laadale.
						Given	Loomad viiakse lauta.
						Given	Loomad viiakse randa.
		11	noored			New	Noored viiakse turule.

					Given	Noored viiakse laadale.
					Given	Noored viiakse teatrisse.
					Given	Noored viiakse randa.
	12	tiimid			New	Tiimid viiakse turule.
					Given	Tiimid viiakse laadale.
					Given	Tiimid viiakse konverentsile.
					Given	Tiimid viiakse randa.
	13	giidid	Subject NP	SVA	New	Giidid vahivad turul.
					Given	Giidid vahivad laadal.
					Given	Giidid vahivad hotellis.
					Given	Giidid vahivad rannas.
	14	härrad			New	Härrad vahivad turul.
					Given	Härrad vahivad laadal.
					Given	Härrad vahivad kirikus.
					Given	Härrad vahivad rannas.
	15	siilid			New	Siilid vahivad turul.
					Given	Siilid vahivad laadal.
					Given	Siilid vahivad metsas.
					Given	Siilid vahivad rannas.
	16	vaalad			New	Vaalad vahivad turul.
					Given	Vaalad vahivad laadal.
					Given	Vaalad vahivad näitusel.
					Given	Vaalad vahivad rannas.
	17	X		AVO	New	Turule viiakse naised.
					New	Laadale viiakse piloodid.
					New	Saali viiakse dotsendid.
					New	Randa viiakse koorid.
	18	X			New	Turule viiakse lambad.
					New	Laadale viiakse piloodid.
					New	Lauta viiakse mutid.
					New	Randa viiakse loomad.
	19	X			New	Turule viiakse lapsed.
					New	Laadale viiakse papid.
					New	Teatrisse viiakse juhid.
					New	Randa viiakse noored.
	20	X			New	Turule viiakse poisid.
					New	Laadale viiakse pastorid.
					New	Konverentsile viiakse direktorid.
					New	Randa viiakse tiimid.
	21	X		SVA	New	Kasvatajad vahivad turul.
					New	Prostituudid vahivad laadal.
					New	Poisid vahivad hotellis.
					New	Giidid vahivad rannas.
	22	X			New	Prouad vahivad turul.

						New	Oivikud vahivad laadal.
						New	Kirjanikud vahivad kirikus.
						New	Härrad vahivad rannas.
		23	X			New	Koerad vahivad turul.
						New	Lehmad vahivad laadal.
						New	Kanad vahivad metsas.
						New	Siilid vahivad rannas.
		24	X			New	Kassid vahivad turul.
						New	Pullid vahivad laadal.
						New	Pastorid vahivad näitusel.
						New	Vaalad vahivad rannas.
		25	X			New	Turul vahivad kassid.
						New	Laadal vahivad pullid.
						New	Näitusel vahivad pastorid.
						New	Rannas vahivad vaalad.
2		26	koorid	Object NP	AVO	New	Turule viiakse koorid.
						Given	Laadale viiakse koorid.
						Given	Saali viiakse koorid.
						Given	Randa viiakse koorid.
		27	loomad			New	Turule viiakse loomad.
						Given	Laadale viiakse loomad.
						Given	Lauta viiakse loomad.
						Given	Randa viiakse loomad.
		28	noored			New	Turule viiakse noored.
						Given	Laadale viiakse noored.
						Given	Teatrisse viiakse noored.
						Given	Randa viiakse noored.
		29	tiimid			New	Turule viiakse tiimid.
						Given	Laadale viiakse tiimid.
						Given	Konverentsile viiakse tiimid.
						Given	Randa viiakse tiimid.
		30	giidid	Subject NP	AVS	New	Turul vahivad giidid.
						Given	Laadal vahivad giidid.
						Given	Hotellis vahivad giidid.
						Given	Rannas vahivad giidid.
		31	härrad			New	Turul vahivad härrad.
						Given	Laadal vahivad härrad.
						Given	Kirikus vahivad härrad.
						Given	Rannas vahivad härrad.
		32	siilid			New	Turul vahivad siilid.
						Given	Laadal vahivad siilid.
						Given	Metsas vahivad siilid.
						Given	Rannas vahivad siilid.
		33	vaalad			New	Turul vahivad vaalad.

					Given	Laadal vahivad vaalad.
					Given	Näitusel vahivad vaalad.
					Given	Rannas vahivad vaalad.
	34	giidid	Object NP	OVA	New	Giidid viiakse turule.
					Given	Giidid viiakse laadale.
					Given	Giidid viiakse hotelli.
					Given	Giidid viiakse randa.
	35	härrad			New	Härrad viiakse turule.
					Given	Härrad viiakse laadale.
					Given	Härrad viiakse kirikusse.
					Given	Härrad viiakse randa.
	36	siilid			New	Siilid viiakse turule.
					Given	Siilid viiakse laadale.
					Given	Siilid viiakse metsa.
					Given	Siilid viiakse randa.
	37	vaalad			New	Vaalad viiakse turule.
					Given	Vaalad viiakse laadale.
					Given	Vaalad viiakse näitusele.
					Given	Vaalad viiakse randa.
	38	koorid	Subject NP	SVA	New	Koorid vahivad turul.
					Given	Koorid vahivad laadal.
					Given	Koorid vahivad saalis.
					Given	Koorid vahivad rannas.
	39	loomad			New	Loomad vahivad turul.
					Given	Loomad vahivad laadal.
					Given	Loomad vahivad laudas.
					Given	Loomad vahivad rannas.
	40	noored			New	Noored vahivad turul.
					Given	Noored vahivad laadal.
					Given	Noored vahivad teatris.
					Given	Noored vahivad rannas.
	41	tiimid			New	Tiimid vahivad turul.
					Given	Tiimid vahivad laadal.
					Given	Tiimid vahivad konverentsil.
					Given	Tiimid vahivad rannas.
	42	X		AVS	New	Turul vahivad naised.
					New	Laadal vahivad piloodid.
					New	Konverentsil vahivad dotsendid.
					New	Rannas vahivad tiimid.
	43	X			New	Turul vahivad õpilased.
					New	Laadal vahivad kardinalid.
					New	Hotellis vahivad professorid.
					New	Rannas vahivad giidid.

		44	X			New	Turul vahivad kasvatajad.
						New	Laadal vahivad prostituudid.
						New	Kirikus vahivad poisid.
						New	Rannas vahivad härrad.
		45	X			New	Turul vahivad koerad.
						New	Laadal vahivad lehmad.
						New	Metsas vahivad kanad.
						New	Rannas vahivad siilid.
		46	X		OVA	New	Juhid viiakse turule.
						New	Papid viiakse laadale.
						New	Lapsed viiakse saali.
						New	Koorid viiakse randa.
		47	X			New	Lambad viiakse turule.
						New	Piloodid viiakse laadale.
						New	Mutid viiakse lauta.
						New	Loomad viiakse randa.
		48	X			New	Poisid viiakse turule.
						New	Pastorid viiakse laadale.
						New	Emmed viiakse teatrisse.
						New	Noored viiakse randa.
		49	X			New	Õpilased viiakse turule.
						New	Kardinalid viiakse laadale.
						New	Professorid viiakse konverentsile.
						New	Tiimid viiakse randa.
		50	X			New	Prouad viiakse turule.
						New	Oivikud viiakse laadale.
						New	Kirjanikud viiakse hotelli.
						New	Giidid viiakse randa.

Appendix B: Additional Figures

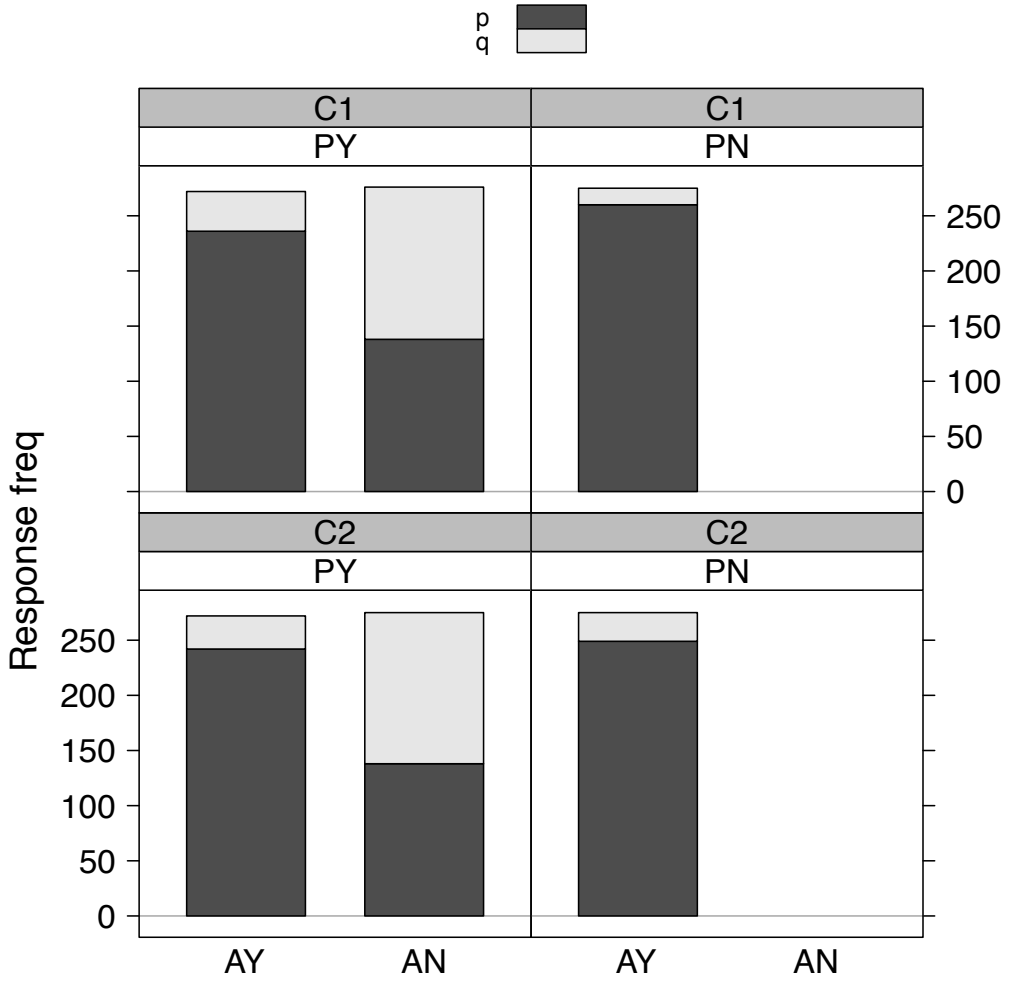

Figure B.6.1. Response frequency in relation to the factors position (P) and accent (A). The factor levels yes (Y) or no (N) encode whether the stimuli presented in the pair differed in the factor or not. C1 means the context that focuses on object, C2 focuses on adverbial. p means that the expected response in the condition is the stimulus-sentence that has pitch accent on the word in focus.

Linear mixed models with dependent variable response frequency (p/q-distribution), with fixed factors context (C1, C2), accent (with difference (Y) or without difference (N)) and position (with difference (Y) or without difference (N)) and with random factors subject and item showed a significant interaction between the fixed factors ($\chi 2[5]=394$, $p<0.001$). The post-hoc Tukey tests showed no significant difference

between the two contexts in any condition, therefore the data from two contexts was pooled together and the new model was defined with fixed factors accent and position.

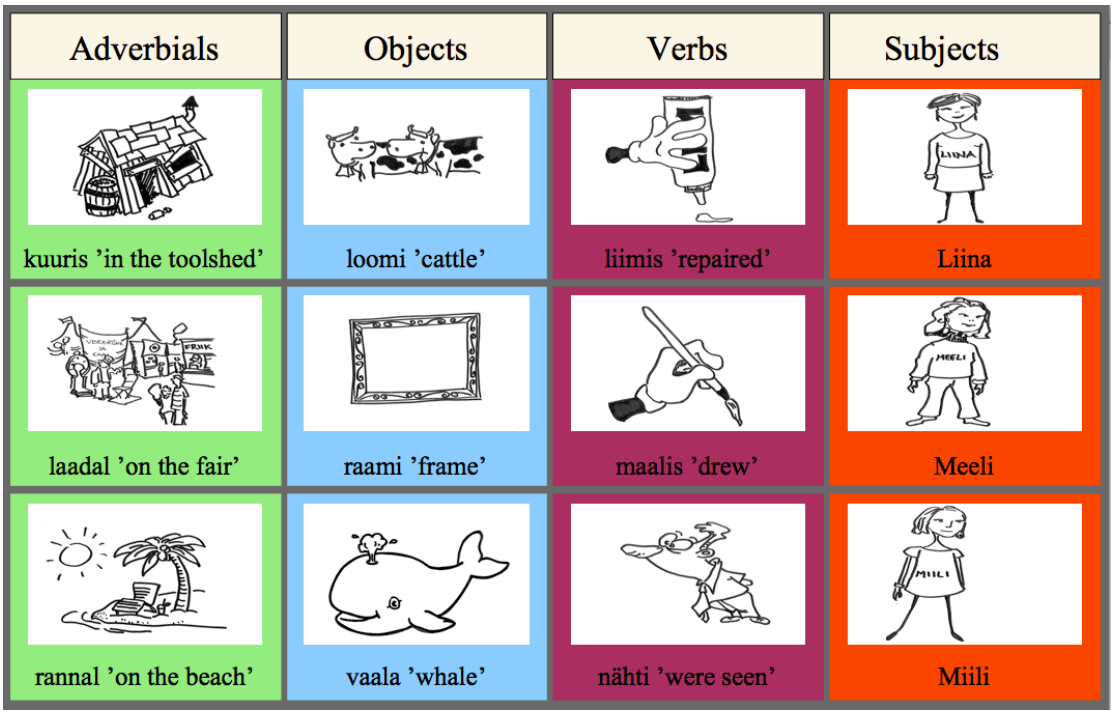

Figure B.6.2. Sample of the pictures together with word forms that were presented to the participants. To ease the processing of the word order, each grammatical function was assigned a different color (subject was orange, verb dark pink, object blue and adverbial green). Every picture appeared within a frame in the corresponding color.

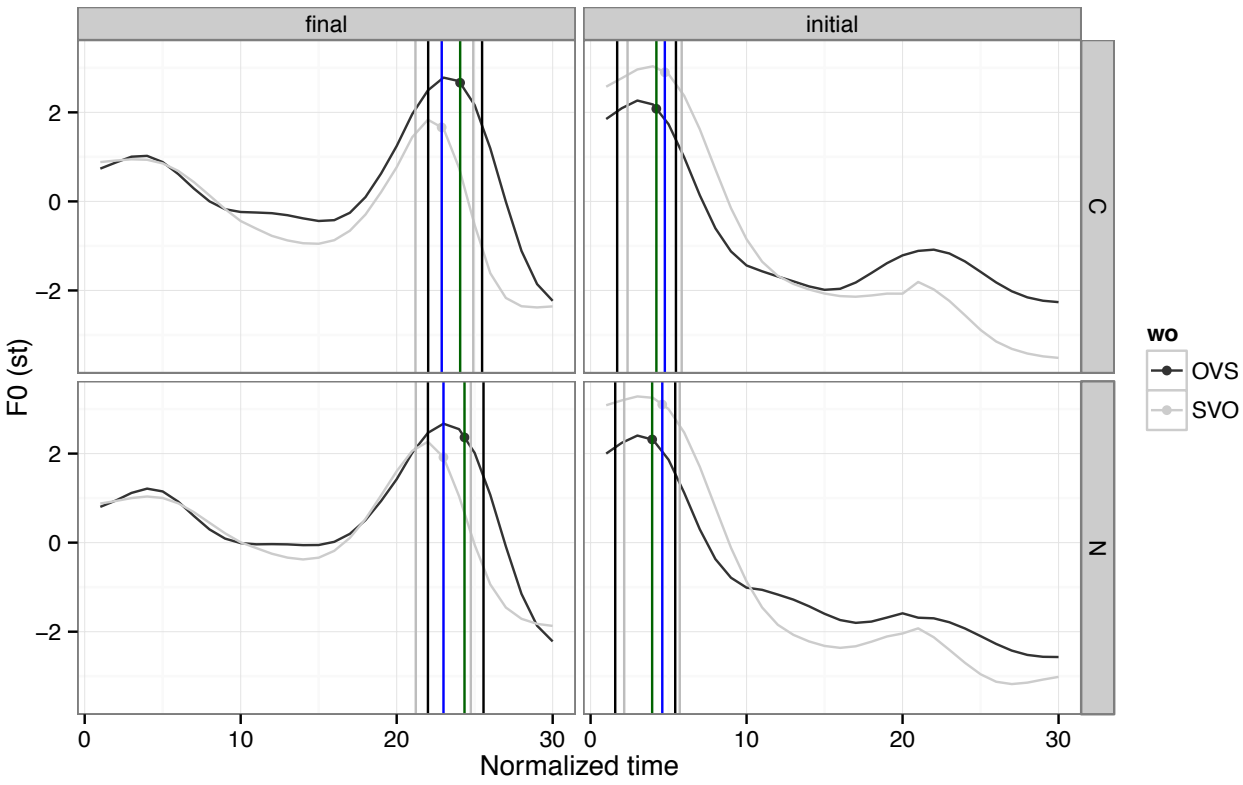

Figure B.6.3. F0 contour aggregated over utterances as a function of normalized time (10 F0-values for each word in the sentence). Initial refers to the utterances where the focus was at the beginning of the sentence. Final refers to the utterances where the focus was at the end of the sentence. *Blue vertical line* is the location of the peak in utterances with SVO word order. *Green vertical line* is the location of the peak in utterances with OVS word order. *Grey black vertical lines* mark the boundaries of stressed vowel in utterances with SVO and OVS word order respectively.

Figure B.3 demonstrates the average location of the peak (vertical blue or green line) within the stressed vowel (vertical grey and black lines) of the word in focus on the background of the pitch contours. It can be observed that across all the experimental conditions the peak in sentence-initial position is located closer to the end of the vowel, whereas in sentence-final position it is approximately in the middle of the vowel and further away from the end of the vowel.