

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is an author's version which may differ from the publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/74861>

Please be advised that this information was generated on 2016-05-02 and may be subject to change.

**Studies on intonation and information structure
in child and adult German**

© 2010, Laura de Ruiter
Cover drawing: Hanna Birmans
Cover design: Ponsen & Looijen, Wageningen
Printed and bound by Ponsen & Looijen, Wageningen

Studies on intonation and information structure in child and adult German

een wetenschappelijke proeve
op het gebied van de Letteren

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus **prof. mr. S.C.J. Kortmann**
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op **vrijdag 5 maart 2010**
om **12.00 uur** precies

door

Laura Edwina de Ruiter

geboren op 1 april 1980
te Berlijn, Duitsland

PROMOTOR: Prof. dr. W. Klein
COPROMOTORES: Dr. B. Braun (Universiteit Konstanz, Duitsland)
Dr. A. Chen (MPI, Nijmegen)

MANUSCRIPTCOMMISSIE:
Prof. dr. P. Fikkert
Prof. dr. A. Cutler (MPI, Nijmegen)
Prof. dr. D.R. Ladd (Universiteit Edinburgh, Verenigd Koninkrijk)

The research reported in this thesis was supported by a grant from the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, München, Germany.

Contents

General introduction..... 9

Part I: Background..... 15

1 Introduction..... 17

 1.1 Reference and information structure..... 17

 1.2 Linguistic markings of information-structural categories..... 24

 1.2.1 *Morphosyntax* 25

 1.2.2 *Intonation* 27

 1.3 The acquisition of linguistic markings of information structure 50

 1.3.1 *Assessment of others’ perspective* 51

 1.3.2 *Morphosyntax* 53

 1.3.3 *Intonation* 56

 1.4 Research questions..... 65

2 Design of the studies and outline of the dissertation 67

 2.1 The corpus..... 67

 2.2 Outline..... 71

Part II: Empirical Studies..... 75

Introduction 77

3 The prosodic marking of topical referents in the “Vorfeld” 79

 3.1 Introduction..... 79

 3.2 Background..... 80

 3.2.1 *Topical referents and contrast* 80

 3.2.2 *Topic marking in adult language*..... 83

 3.3 Topics in child language 86

 3.4 Method 88

 3.5 Item selection and annotation..... 90

 3.5.1 *Item selection* 90

 3.5.2 *Annotation*..... 91

 3.5.3 *Consistency check of accent labels*..... 93

 3.6 Results and discussion 97

 3.6.1 *Phonological analysis – choice of pitch accent type*..... 97

 3.6.2 *Phonetic analysis* 101

Contents

3.7	General Discussion	103
4	The marking of information status in narrative discourse	109
4.1	Introduction.....	109
4.2	Method.....	111
4.2.1	<i>Participants.....</i>	<i>111</i>
4.2.2	<i>Materials and design.....</i>	<i>112</i>
4.2.3	<i>Data pre-processing and selection.....</i>	<i>117</i>
4.3	Prosodic annotation and consistency check of accent labels.....	125
4.4	Results.....	126
4.4.1	<i>Boundary tones</i>	<i>126</i>
4.4.2	<i>Pitch accents.....</i>	<i>128</i>
4.5	Within-accent comparison across age groups.....	136
4.5.1	<i>Analysis and results.....</i>	<i>137</i>
4.6	Discussion.....	141
5	The marking of identifiability and activation	153
5.1	Introduction.....	153
5.2	Participants and materials	155
5.3	Procedure, item selection and annotation	156
5.4	Results and discussion.....	156
Part III: Methodological studies		161
Introduction		163
6	Information status marking in read speech	165
6.1	Introduction.....	165
6.2	Materials and procedure.....	165
6.3	Participants.....	166
6.4	Data annotation, data selection and accent label consistency check.....	166
6.5	Results and discussion.....	167
7	Naïve listeners' perception of prosodic boundaries	173
7.1	Introduction.....	173
7.2	Method.....	177
7.3	Materials	177
7.4	Participants.....	179
7.5	Analyses and results	179

7.5.1	<i>Reliability analysis</i>	179
7.5.2	<i>Boundary scores</i>	181
7.6	Discussion.....	182
8	Polynomial modeling of spontaneous child and adult intonation ..	185
8.1	Introduction.....	185
8.2	Method.....	187
8.2.1	<i>Data and annotation</i>	187
8.2.2	<i>Polynomial modeling</i>	190
8.3	Results.....	196
8.4	Extending the model: Adding alignment parameters	209
8.5	Investigating age-related differences with polynomials.....	213
8.6	General discussion.....	217
9	Summary and conclusions	223
	References	231
	Appendices	251
A	Picture stories	251
B	Pitch accent label consistency check.....	260
C	Consistency check of pitch accent labels by age groups	282
D	Statistical analysis for target words only.....	301
E	Within-accent category comparison across age groups	302
F	Texts used in the reading experiment.....	309
G	Pitch accent label consistency check for the reading experiment.....	312
H	Four-coefficient profiles for nuclear contours	314
	Samenvatting	315
	Curriculum Vitae	325

General introduction

When we speak, we talk about all kinds of things, such as

- persons: *Gandhi, Harry Potter, my mother, she, that one;*
- places: *Paris, the dark side of the moon, under the sofa;*
- animals: *Charlie, a unicorn, your goldfish, it;*
- objects: *the Eiffel Tower, a chair, the knife in the bottom drawer, those ones over there;*
- events: *9/11, the fall of the Roman Empire, yesterday's flood warning, it;*

and many others. Language allows us to talk about things that are not present in the here and now. We have mental representations of them, and when we engage in discourse, we refer to those mental representations. However, when making a contribution to the discourse, speakers have several options to encode a referent linguistically. Typically, these include different kinds of noun phrases, which vary in their morphosyntactic shape and richness of lexical content they carry, such as:

- names, such as *Harry Potter, Elvis, Gandhi;*
- definite lexical noun phrases, such as *the dark side of the moon, the knife in the bottom drawer;*
- pronouns and similar forms, such as, *I, he, it, they, those;*
- "empty" elements, such as the elided second subject in *John tried hard but \emptyset failed.*

The speaker's choice to use one form and not another depends on a number of factors. The most important one is the amount of contextual information that is already available in the given communicative situation. Definite noun phrases, for instance, usually presuppose that the entity referred to is somehow "familiar" to the interlocutor (e.g., *the Eiffel tower*), whereas this is not the case for indefinite noun phrases. Some types of pronouns normally require that the entity referred to has already been mentioned in the preceding discourse (*he, they*), or that it is identifiable by its communicative role in the situation (*I, you*).

General introduction

The morphosyntactic properties of these different types of noun phrases and the conditions that determine or influence their choice in a given communicative situation have been comparatively well studied for a number of languages. This is much less true for a different type of variation in the properties of referring expressions: the way in which they are *prosodically* realized. If, for example, a referent is somehow “given”, because it has been mentioned before, or because the situation is such that the interlocutors have it in their focus of attention, then this not only restricts the type of referring expression that can be used, but it also influences the intonation of that expression. One basic assumption is that a given referent is “deaccented”, as in *On the right you’ll see a CHURCH. Walk PAST that church...*, where the referent *church* is deaccented on its second mention. The intonational properties of the referring expression may also vary with a particular pragmatic function that it has within the sentence. For instance, a sentence is often assumed to have a “topic” (what the sentence is about) and a “comment” (what is being said about the topic). These two functions are marked by, among other things, intonation.

Factors of this sort, which influence a speaker’s choice of a referring expression and its prosodic realization, and how listeners interpret a given utterance, belong to what is known as *information structure* (IS). Thus, what is of interest in the study of IS is not the propositional content of utterances (the “what”), but the different ways this content can be encoded (the “how”), and the factors that determine the choice of those alternatives.

Information structure has fascinated linguists from as early as the 19th century (e.g., Paul, 1880; Weil, 1844), and this has resulted in an enormous wealth of theories and publications. No contemporary work on information structure, therefore, comes without a warning of the conceptual and terminological jungle one is about to enter, impressively illustrated in a diagram by Kruijff-Korbayová and Steedman (2003), reproduced in Figure 1.

General introduction

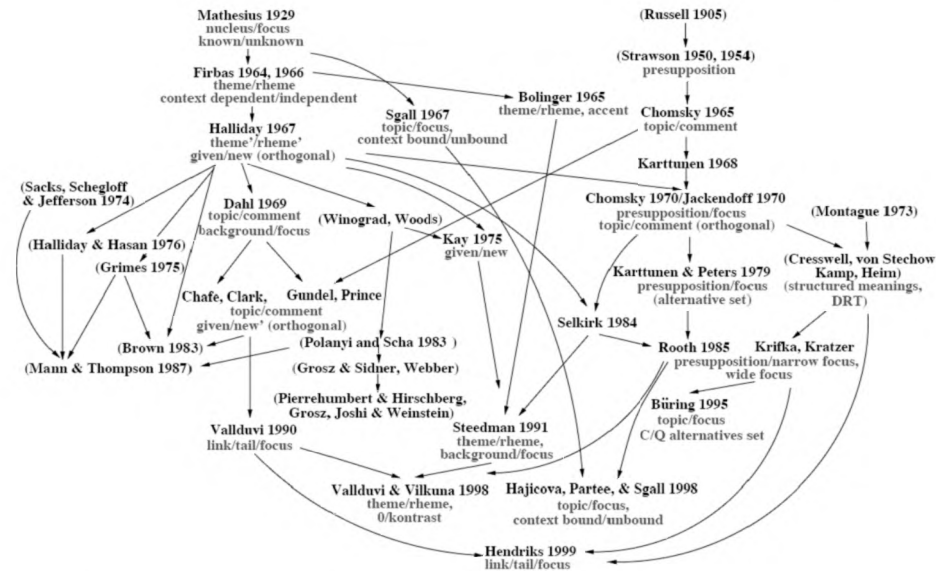


Figure 1: The jungle of information structure. From Kruijff-Korbayová and Steedman (2003, p. 260), reprinted with permission.

The complexity is not only terminological or conceptual; it also reflects the wealth of linguistic means that are used to express distinctions in information structure. These include the choice that speakers make among the types of referring expressions, the intonation of these and of other elements in the sentence, but also word order, the use of ellipsis, as well as special particles such as the Japanese *wa* and *ga* (which express topichood and subjecthood, respectively).

The complexity of information structure is not only daunting for the adult language user, but even more so for children who have to master it during language acquisition. Not only does a child have to learn that *Hund* 'dog' has masculine gender in German, for example, but also when to refer to such an animal with a definite or indefinite expression, and how to vary the intonation of this expression. Children's use of referring expressions has received quite a lot of attention; we know for instance that even six-year-old children tend to introduce new referents inappropriately with definite noun phrases (see e.g.,

General introduction

Hickmann, 2003). In contrast, our knowledge about children's use of intonation for information structure marking is very limited. Previous studies have been based either on anecdotal evidence, or on production in rather unnatural contexts, such as in imitation tasks. A more comprehensive study investigating these issues in children's *spontaneous* speech production is therefore lacking. Furthermore, typically no distinction has been made between the *phonological* and *phonetic* aspects of intonation. However, in order to become competent speakers, children have to master both: On the one hand, they have to learn the form-meaning relationship, such as 'the topic of a sentence may be marked by a rising intonation contour'. On the other hand, they also need to learn how exactly a given contour is actually physically realized, similar to the way a child has to learn the language-specific way in which a certain consonant is pronounced.

The overall aim of this dissertation is to get a better picture of how children use intonation to mark information structure in natural discourse, taking into account both the phonological and the phonetic level. More specifically, we shall look at these two questions:

- How do topic-hood and givenness influence the intonational realization of referring expressions?
- How do children and adults differ in this regard?

When investigating these questions, there are also some methodological issues to be considered. For example: What should actually be regarded as the target model for children? Studies on information structure marking in adult intonation typically employ reading tasks, but to what extent are those intonation patterns comparable to what children hear in everyday communication? Therefore, this dissertation is organized into two parts: an empirical one, which focuses on the two questions raised above, and a methodological one, which is concerned with more general issues of intonation research and research into child intonation in particular.

The first two chapters provide a theoretical background for the dissertation. Chapter 1 first introduces relevant concepts in information structure and intonation (sections 1.1 and 1.2). Then the current state of research regarding

General introduction

children's linguistic marking of information structure is reviewed (section 1.3). This leads us to the specific research questions that are being addressed in this dissertation, which are formulated in more detail in section 1.4. Chapter 2 describes the type of speech data that has been collected to answer those questions (2.1), and outlines the individual studies presented in the remainder of the dissertation (section 2.2).

Part I: Background

1 Introduction

1.1 *Reference and information structure*

As stated in the general introduction, this thesis is concerned with aspects of information structure (IS). This section provides an introduction to the basic concepts and notions of IS that are needed in order to understand their operationalization in the empirical parts of this thesis. More comprehensive introductions and discussions can among others be found in Vallduví and Engdahl (1996), von Heusinger (1999), Kruiff (2001) or Kruijff-Korbayová and Steedman (2003). The framework I am adopting is to a large extent based on the influential work of Lambrecht (1994), Chafe (1974, 1987, 1994), and Prince (1981).

Underlying most theories of IS is the assumption that speakers organize their utterances into an informative part and a less informative one. Different terms have been suggested to capture this dichotomy, such as given/new or theme-rheme (for an overview, see e.g., von Heusinger, 1999, p. 102ff., and Figure 1 above). However, a number of scholars (Gundel, 1999; Halliday, 1967a; Lambrecht, 1994) have noted that these terms and concepts conflate two different information-structural categories. The first category concerns the mental representations of the discourse referents. In the interlocutors' minds, these referents can be either *given*, for example because they have just been mentioned, or they can be *new*. Thus this first dimension is about the cognitive *status* (a term introduced by Gundel, Hedberg, & Zacharski, 1993) of referents.

The second category concerns the pragmatic *functions* that these referents can fulfill in a sentence. One such pragmatic function is topic-hood. A referent can function as the topic of a sentence, in which case the sentence is interpreted as conveying information about that referent. These two notions are illustrated in (1). In this and all subsequent examples, the referent in question is underlined.

1.1 Reference and information structure

- (1) I bought an interesting book yesterday. It's on information structure.

In (1), *an interesting book* is a new referent the speaker introduces to the discourse. It is subsequently given in the discourse (as a consequence referred to by the personal pronoun *it*). In addition to being given, the referent functions as a *topic* in the second sentence. The sentence says something about the referent (namely that it is on information structure). Similarly, a referent can have the role of *focus*, as in (2).

- (2) A: What did Jenny order?
B: (She ordered) a pizza.

The referent *a pizza* is the focus of the utterance. In Lambrecht's account (but also in e.g., Jackendoff, 1972), the focus is what turns an open proposition (here: "Jenny ordered x") into a proper assertion (*Jenny ordered a pizza*).¹

Example (2) illustrates another important difference between the status of referents and their pragmatic functions. While notions like topic and focus apply to the sentence or utterance level, the status of referents needs to be assessed with respect to the discourse level.² The *interesting book* in (1), for instance, is now an established referent in the universe of discourse, is part of the *common ground* (Clark, Schreuder, & Buttrick, 1983; Stalnaker, 1978), and will stay so – irrespective of the pragmatic function it may assume in any subsequent utterance.

Yet, although the discourse participants now "know" that this book exists, their awareness of it is likely to vary. As the conversation proceeds, other ref-

¹ It should be mentioned at this point that "focus" is at times also used as a label for an element of the utterance that is linguistically marked, for example by an accent (as being important). In these cases, focus is used merely as a synonym for prominence. Focus in that sense may be coextensive with the informative component, but it need not be.

² Note that I am excluding the notion of *discourse topic* here, which concerns the larger "theme" of a longer stretch of speech such as the paragraph or an entire story (e.g., Givón, 1976).

1.1 Reference and information structure

erents will be talked about, and the attention will move away from the interesting book – until it can be brought up again at a later point. This distinction between speakers' and hearers' *knowledge* of (the existence of) a discourse referent on the one hand, and their *awareness* or *consciousness* of it on the other has been referred to as the difference between *identifiability* of a discourse referent and its *activation* (Lambrecht, 1994).

An identifiable referent is a referent of which the speaker assumes the hearer to already have a mental representation, or a metaphorical "file-card" (Heim, 1982; Vallduví, 1992). Consequently, the hearer should be able to pick out and identify the referent the speaker has in mind from all referents that could be designated with a particular linguistic expression (Lambrecht, 1994, p. 77). In (1) above, the hearer is likely to identify the referent of *it* as the interesting book the speaker has just introduced. A referent can also be identifiable because the interlocutors have mutual knowledge of this referent prior to the discourse situation (e.g., *the sun*, or *the stupid wagon-wheel coffee table*), or because it is saliently present in the external discourse situation, as in (3).

- (3) [Ann trying to put up a picture on the wall, saying to Joe]
Pass me the hammer, will you?

(Lyons, 1999, p. 6)

The term activation, on the other hand, reflects the view that the status of a referent in discourse correlates with cognitive degrees of consciousness or awareness of this referent in the mind of the interlocutors. A person has knowledge of countless referents, but only a limited number of them will be in the centre of his or her attention at any given point in time. As Lambrecht (1994) puts it: "Knowing something and thinking of something are two different things" (p. 93).

In contrast to identifiability, which is a binary variable (a referent is either identifiable or not), activation is usually more graded. Scholars differ in whether they postulate a limited set of concrete activation states (e.g., Prince' 1981 assumed familiarity scale) or a continuum (e.g., Lambrecht, 1994). For the remainder of this thesis, I will follow Chafe (1987) and Lambrecht (1994),

1.1 Reference and information structure

who both postulate at least three activation states: inactive, semi-active, and active. Chafe describes these three activation states as follows:

An active concept is one that is currently lit up, a concept in a person's focus of consciousness. A semi-active concept is one that is in a person's peripheral consciousness, a concept of which a person has a background awareness, but which is not being directly focused on. An inactive concept is one that is currently in a person's long-term memory, neither focally nor peripherally active.

(Chafe 1987, p. 25)

With activation being the central concept of Chafe's notion of givenness, he defines givenness of referents in terms of the *activation cost* that a speaker has to invest in order to promote a referent to the active state in the mind of the listener. If a referent is already "lit up" in the listener's mind, it is *given*, and not much effort is needed to maintain it that way. If the referent is transferred to the active state from a previously inactive state, it is *accessible*. Finally, if the referent becomes activated from a previously inactive state, it is *new*. These ideas are illustrated in Figure 2.

1.1 Reference and information structure

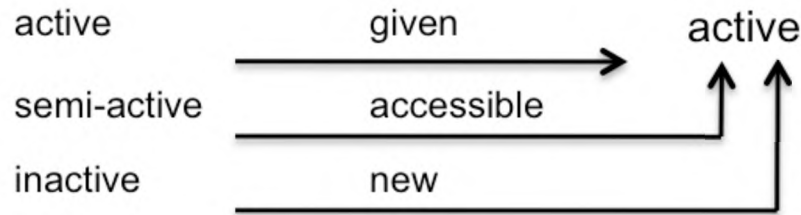


Figure 2: Givenness and activation in Chafe's framework (from Chafe, 1994, p. 73).

It is generally assumed that a referent is brought into a person's focus of consciousness, or is activated, by being mentioned in the immediate context. This is also the operationalization of this concept that will be used in this thesis. As for the semi-active status, a referent can acquire this status for three different reasons, according to Chafe and Lambrecht: First, it may become deactivated from the active state (e.g., due to other intervening referents). Second, it may be promoted from the inactive state via inference, such as a semantic relationship with a preceding referent (e.g., part-whole relations like *handbrake* – *car*), or because it is part of a larger concept or “semantic frame” (Fillmore, 1982). Third, a referent may become semi-active because of its presence in the extralinguistic setting.

Lambrecht's system of givenness relations implies that a referent needs to be identifiable in order to have one of the three activation states (Lambrecht 1994, p. 109). This is logically true, since a representation of a referent can only have some degree of activation if there is a representation in the first place. Still, as observed by Prince (1981), non-identifiable referents (“brand-new” in her terminology) and identifiable yet hitherto “unused” referents (also Prince's term) are both in a way new to the listener. In the schematic representation of the “system of givenness” terms in Figure 3 below this commonality is indicated by the bracket subsuming both brand-new and unused referents under the label *new*.

There is of course no way of objectively measuring a referent's degree of activation in someone's mind. The speaker can only make assumptions about

1.1 Reference and information structure

the activation of referents in the listener's mind. This means that whenever we use terms such as "activation state", they have to be implicitly qualified as assumed. In the following chapters, I will for the most part use the labels new, given, and accessible instead of inactive, active and semi-active to avoid the connotation that a referent's degree of activation can be observed from the outside. These are also the labels used by Baumann (2006) and Baumann and Grice (2006), whose work on the marking of givenness in German is particularly relevant for the investigations presented in this thesis, as we shall see in section 1.2.2.2 below.

1.1 Reference and information structure

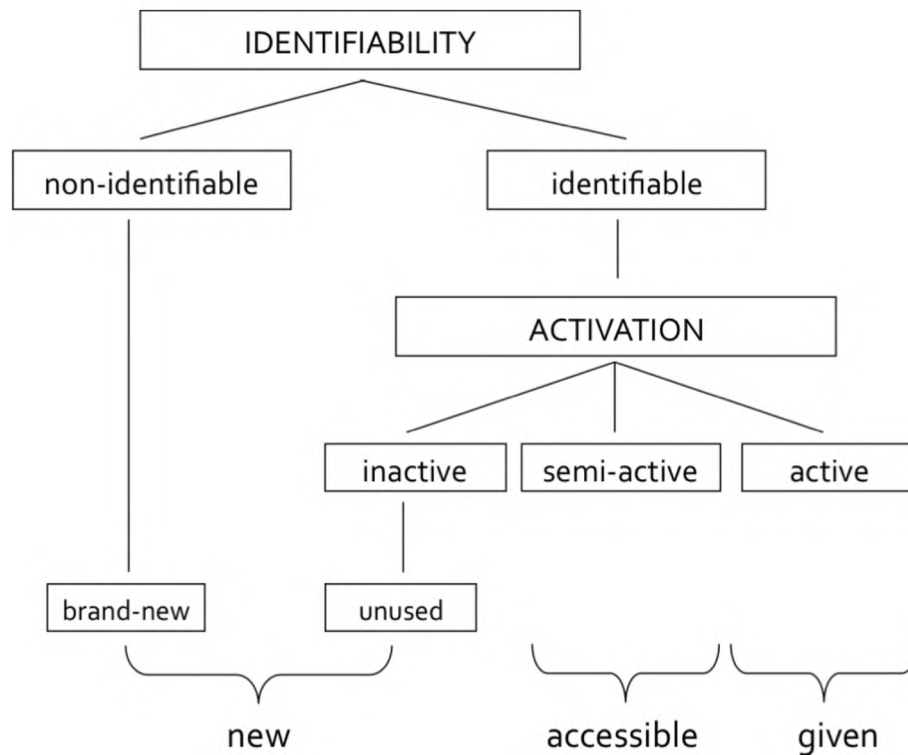


Figure 3: Taxonomy of givenness relations (based Baumann, 2005, p. 76, following Prince, 1981 and Lambrecht, 1994).

As briefly mentioned above, the information status of a referent (i.e., its identifiability and activation) on the one hand, and its pragmatic function in a sentence on the other hand are in theory independent of each other. But there are natural correlations between the two. Most topic referents are identifiable, which is not surprising if we assume that by making a referent the topic of a sentence, the speaker wishes to add something to the hearer's knowledge of that referent: In most cases, this is only sensible if the hearer already has a mental representation that this new information can be linked up to. In addition to being identifiable, topics of sentences are often active or semi-active

1.2 Linguistic markings of information-structural categories

referents. Lambrecht (1994, p. 160ff.) argues that this has to do with the increased processing effort that would arise if the hearer were presented with a less activated referent (e.g., an unused one) as the topic. The processing of the proposition that expresses new information about the topic referent would have to be carried out simultaneously with the process of retrieving or inferring the referent itself.

Note that Lambrecht's hypothesis that speakers do not normally "burden" the hearer with utterances that would be difficult to process implies that speakers "design" their contributions for the benefit of the hearer. The idea that speakers make assumptions about the addressee's knowledge and state of consciousness³ at the time of the utterance, and tailor their utterances accordingly is a key tenet of cognitively-oriented approaches to IS such as the one advocated by Lambrecht (but see Barr & Keysar, 2005 for a different view). In this thesis, I adopt the general view that speakers make assumptions about the addressee's state of mind, but for the moment I leave open to which degree this influences their utterances.

Equipped with the necessary background knowledge of the information-structural concepts of information status and pragmatic function, we are now ready to take a look at their actual realization in language, which is the subject of the following section.

1.2 *Linguistic markings of information-structural categories*

There are three major ways in which the information status of referents and their pragmatic functions are reflected in language: morphosyntactic form of the referring expression itself, intonation, and syntactic form of the utterance (i.e., its word order and whether it constitutes a specific syntactic construction). This section provides a survey of the possible reflections of IS in the form of referring expressions and the intonation of this expression (for effects of word order or syntactic constructions, see e.g., Erteschik-Shir, 2007; Lambrecht, 1994).

³ This perhaps slightly misleading term refers to the extent to which the addressee is aware of or paying attention to the things being talked about.

1.2 Linguistic markings of information-structural categories

1.2.1 Morphosyntax

In the morphosyntactic domain, many languages, including English and German, signal the difference between identifiable and unidentifiable referents via the distinction between definite and indefinite noun phrases, as in (1) and (3) above, repeated here as (4) and (5).⁴

- (4) I bought an interesting book yesterday.
(5) Pass me the hammer, will you?

However, nominal determiners have other functions than identifiability marking as well (discussed in Hickmann, 2003), such as the marking of specific and non-specific reference. Indefinite noun phrases can also be used to indicate a particular type or class of referents. Compare the sentences in (6):

- (6) a. I'd like to have a car. [non-specific]
b. I'd like to have the car. [specific]

In (6a), the speaker is indicating that she would like to own an object from the general class of cars, whereas in (6b) she has a particular instantiation of that class in mind. Similarly, indefinite determiners can be used in so-called naming constructions, in which the speaker is assigning a class label to a referent. In (7), the referent in question is referred to with a demonstrative pronoun (*that*), while the indefinite form is part of the predicate, which informs us about some property of the referent.

⁴ A number of scholars have pointed out that the correlation between (un-)identifiability and (in-)definiteness is far from perfect, and have suggested other semantic and pragmatic criteria that licence definiteness instead, such as inclusiveness (Hawkins, 1978; Lyons, 1999) or salience (Lewis, 1979). The details of this theoretical debate are not pertinent to the studies presented in this thesis, and won't be discussed further. The interested reader is referred to Lyons (1999), who provides a comprehensive comparative and theoretical discussion of the issues involved.

1.2 Linguistic markings of information-structural categories

- (7) [Child pointing to a cat running past]
That's a cat!

Example (7) also illustrates another important distinction that has not been introduced so far. The speaker uses the demonstrative *that* to “point” to a referent in the real world. This distinction has been called that between exophoric and endophoric reference⁵ (Halliday & Hasan, 1976, comparable to what Lambrecht calls “situationally accessible” and “textually accessible”), and appears to play an important role in language acquisition. Exophoric references point to a referent in the immediate non-linguistic situation, whereas endophoric references relate to a referent in the linguistic context. The demonstrative in (7) is a prototypical example of exophoric reference. In contrast, if (8b) was uttered by the addressee of (1) above (repeated here as 8a) at a later point in the discourse, *the book* would indicate a referent in the linguistic context, and would be an example of endophoric reference.

- (8) a. I bought an interesting book yesterday. It's on information structure
b. Could you lend me the book you've mentioned?

Note that the morphosyntactic marking of a referent's status as identifiable is the same, irrespective of the source of this givenness (i.e., it does not matter whether its identifiability comes from the text-internal or the text-external world). However, in the developmental studies to be discussed later (section 1.3.2), we shall find that the endophoric/exophoric distinction appears to be a particularly relevant one.⁶

⁵ Exophoric reference is often also referred to as “deixis”, and endophoric reference as “anaphoric reference”.

⁶ For more examples of this distinction, including temporal reference, see Hickmann (2003: 39ff.), among others.

1.2 Linguistic markings of information-structural categories

1.2.2 Intonation

The other important device for marking the information state of a referent – and the one that is central to the studies in this dissertation – is intonation. It is commonly assumed that new referents are realized with a pitch accent (e.g., Brown, 1983; Chafe, 1974; Prince, 1981). In (1), repeated here as (9), *book* is canonically realized with an accent on it. In this and all subsequent examples, capital letters indicate an accent on the word.

(9) I bought an interesting BOOK yesterday.

According to Lambrecht, this holds for both brand-new referents (i.e., non-identifiable) and for unused referents. If an identifiable but hitherto unmentioned referent is brought into the discourse, it too receives an accent (e.g., *Pass me the HAMMER, will you?*). In Chafe's cognitive approach this is explained by assuming that although the hearer may have some background awareness of the referent (e.g., because the hammer is visible), its representation is not activated and needs to be activated by using intonational prominence. Thus the morphosyntactic marking of identifiability and the intonational marking of activation are to some extent independent of each other. In contrast to new referents, active referents (encoded either as full noun phrases or as pronouns) are commonly assumed to be characterized by a lack of prominence, illustrated in (10) and (11), where the referent of *that sweater* and *it*, respectively, are given. As a consequence, in both cases the accent shifts to the next constituent that can be accented (here: *like*).

1.2 Linguistic markings of information-structural categories

- (10) A. Why don't you wear your red sweater?
B. I don't LIKE that sweater.
- (11) A. Why don't you wear your red sweater?
B. I don't LIKE it.

Accessible (or semi-active) referents are typically expressed as full lexical forms, but different proposals have been made regarding their intonational marking. In some accounts (e.g., Lambrecht, 1994) it is simply presumed that a speaker can choose whether to mark accessible referents like new ones (i.e., with accent) or like given ones (i.e., without an accent). Other researchers have suggested that the different activation states of referents are signaled by *types* of accents, such as rising vs. falling (Allerton, 1978; Baumann & Grice, 2006; Pierrehumbert & Hirschberg, 1990). Different accent types have also been proposed for the different pragmatic functions that discourse referents can assume (e.g., Steedman, 2000). This assumption is reflected in terms like “topic accent” or “focus accent”. Compare the following two contexts⁷:

- (12) What about the rest of the class? Have you heard anything?
B. [JILL]_T has a TEACHING job now.
- (13) A. Do you know who got the teaching job?
B. [JILL]_F got it.

In (12), *Jill* is a topical referent, and most likely realized with a rising accent, or at least with an accent that is followed by a sustained high pitch. Importantly, it is also followed by another accent in the phrase (here on *teaching*). In (13), on the other hand, *Jill* is the focus of the sentence, and in this case the pitch is clearly falling (and not followed by any other accent). Categorizing the intonational marking of discourse referents only in terms of “accented” or “deaccented” therefore appears too simplistic, and intonation researchers have

⁷ Square brackets and a subscripted F for focus or T for topic indicate the focus or topic domain, respectively.

1.2 Linguistic markings of information-structural categories

meanwhile examined and described the intonation patterns that occur in speech in a much more detailed way. The next section provides some background on intonation in general and introduces the theoretical framework that I am adopting in this thesis. Subsequently, I will describe the annotation system for German intonation used here, and also briefly discuss some methodological problems in connection with the annotation of intonation.

1.2.2.1 Intonational structure

The term *intonation* is used to refer to the (meaningful) movements of the fundamental frequency (F0). F0 is a property of the acoustic signal that depends on the frequency at which the vocal folds vibrate, and it is what we perceive as pitch. Thus, strictly speaking, there is a difference between F0 (the acoustic parameter) and pitch (the perceptual impression), but since the distinction is not crucial for the present investigation, I will use these two terms interchangeably, unless otherwise indicated. The pitch movements serve several functions in language. As shown in the preceding section, they can be used to highlight constituents in an utterance, such as the topic or focus of the sentence, or to mark sentence modality (question vs. statements). In addition to this grammatical or linguistic function, intonation is also used to express attitudes and emotions: Certain intonation patterns can make a person sound angry or bored, or give the impression that a speaker is being sarcastic. These functions of intonation are typically referred to as “paralinguistic”, as they do not contribute to the core semantic meaning of the utterance. However, since they are all expressed by the same cues, the linguistic and paralinguistic functions of intonation interact, which is one factor that makes the investigation of intonation a challenging task, especially in natural speech.

In so-called “intonation languages” such as German and English, pitch movements mark linguistic distinctions on the *utterance level*. This is in contrast with tone languages like Mandarin Chinese, where pitch movements are used to express lexical differences, that is, differences between words that have otherwise the same segmental structure. An often-cited example is the word *ma*, which – depending on the pitch pattern with which it is produced – can mean ‘horse’, ‘mother’, or ‘scold’. Lexical differences in intonation lan-

1.2 Linguistic markings of information-structural categories

guages, on the other hand, are signaled via other acoustic means, which brings us to the difference between intonation and another, closely related term:

prosody.

Prosody is a wider term, encompassing not only intonation but also often other phenomena like word-level stress (e.g., PROject vs. proJECT), rhythm, phrasing (i.e., the chunking of speech) or speech rate. These are typically marked using the acoustic parameters of intensity and duration, which have their perceptual correlates in loudness and length.⁸ (The issue of phrasing will be discussed in more detail in chapter 7.) Like pitch, these parameters also contribute to the perception of syllables as salient. This has led to some confusing terminology in the literature, where terms like “stress” are used by some authors to refer to the word level and by others to the utterance level. In this thesis, I will use *stress* to mean lexical stress, and *accent* to mean utterance-level prominence of the type shown for instance in example (12) above.

Accented syllables not only show a prominence-lending pitch movement, but also differ from unaccented syllables in intensity and duration. This is because lexically stressed syllables are the “docking sites for accent placement” (Sluijter & Van Heuven, 1996, p. 2471). With very few exceptions then (see Shattuck-Hufnagel, Ostendorf, & Ross, 1994), accented syllables are always lexically stressed, too. We shall see that the way in which pitch movements are associated with the lexically stressed syllable play a central role in the currently most widely used framework in intonation, autosegmental-metrical theory, which will be introduced below.

All acoustic parameters that play a role in prosody and intonation – including F0 – are continuous in nature. On the other hand, it was already mentioned that people have proposed different accent *types* as markers of certain information structural categories, indicating that there are *categories* of F0 movement. However, while no one doubts the existence of tonal categories in tone languages (cf. the *ma* example above), the existence of tonal categories in intonation languages has been and is still being debated (e.g., Xu & Xu, 2005). The question of whether or not there is a *phonology* of intonation is the central

⁸ Vowel quality and spectral tilt have been found to play a role as well.

1.2 Linguistic markings of information-structural categories

issue that separates two different approaches to intonation from each other, which, following Ladd (2008), I will call the “phonetic” and the “phonological” approach. I will briefly sketch the differences between the two.⁹

As argued by Ladd (2008), historically, the difference between these two approaches appears to have been methodological in nature. On the one hand, researchers with a background in phonetics were trying to find instrumentally measurable cues in the speech signal that could be linked to certain pragmatic distinctions, such as statements vs. questions. On the other hand, descriptive linguists were interested in finding the general melodic patterns in a given language by impressionistically describing the pitch curves they heard. So the approaches seemed to differ mainly in working instrumentally vs. working impressionistically. However, Ladd points out that a closer look at the two approaches reveals that there is also a more fundamental, theoretical difference. In the first case, there is an underlying assumption of a direct link between phonetic detail (e.g., pitch excursion) and certain pragmatic notions such as “contrast” or “focus”, without a mediating level of abstract representations (e.g., Eady, Cooper, Klouda, Mueller, & Lotts, 1986; Xu, 2005; Xu & Xu, 2005). In contrast, the phonological approach assumes that there is an abstract level of phonological representation, that is, a set of distinct categories, which are then in turn used to express linguistic meaning.

Although the phonological approaches have an abstract level of representation as their common denominator, they vary again along a number of dimensions, such as the nature of the abstract representations (are we talking about tone *levels* or rather tonal *movements*?), or the way in which meaning is ultimately achieved (does it come from the entire contour – the tune – or from the parts that make up the contour – the tones?). Nowadays, the differences between phonetic and phonological approaches are mainly theoretical, but some differences in methodology persist, as not all who look at intonation from a phonological viewpoint use instrumental analysis.

⁹ The difference between phonetic and phonological approaches is discussed in more detail by Ladd (2008), and Arvaniti and Ladd (2009).

1.2 Linguistic markings of information-structural categories

In this thesis, intonation is discussed within what has been come to be known as the autosegmental-metrical (AM) theory of intonation (a term coined by Ladd, 1996), which is a phonological model of intonation. Precursors of the AM intonation model can be found in the 1970's, but it is Pierrehumbert's (1980) doctoral dissertation on American English intonation (see also Beckman & Pierrehumbert, 1986) that is generally seen as the origin of this widely used framework. Since then, AM accounts of intonation have been applied to a variety of other languages (e.g., Gussenhoven, 1984 for Dutch; Sosa, 1999 for Spanish). The framework's dissemination was especially accelerated by the introduction of a standard prosodic annotation system called ToBI (Tones and Break Indices), which I will come back to below. The following paragraphs provide a general introduction to the basic theoretical assumptions and notational conventions of the AM framework.¹⁰

In the AM framework, intonation patterns are described as sequences of high (H) and low (L) tones, which is why it is also referred to as a tone-sequence model. When these tones occur on lexically stressed syllables, they are called *pitch accents*. Pitch accents are conventionally notated with the diacritic "*", and can be either monotonal (i.e., H* or L*) or bitonal (e.g., L*+H, H+L*).¹¹ In the case of bitonal accents, the starred tone indicates which of the two tones is associated with the stressed syllable.¹² The other tone is joined with the starred tone with a "+". The non-starred tone is called a leading tone, if it precedes the starred tone, and a trailing tone if it follows it.¹³ Tones are found not only on stressed syllables, but also at the edges of phrases (i.e.,

¹⁰ For more detailed information about its historical background as well as discussions of some unresolved issues, the reader is referred to Ladd (1996, 2008).

¹¹ This is the case for the original analysis of American English intonation by Pierrehumbert (1980). Subsequent analyses of other languages include tritonal accents as well (e.g., Gussenhoven, 1984 for Dutch; Prieto, D'Imperio, & Fivela Gili, 2005 for Italian).

¹² But see Arvaniti, Mennen and Ladd (2000) for a critical assessment of this interpretation.

¹³ Several AM accounts reject leading tones (Grabe, 1998), thus their accent inventory features only right-headed accents.

1.2 Linguistic markings of information-structural categories

chunks of speech). These edge tones are referred to as *boundary tones*, and marked with the diacritic “%”. They are always monotonal (i.e., H% or L%). Figure 4 illustrates the annotation scheme with two different combinations of pitch accents and boundary tones.

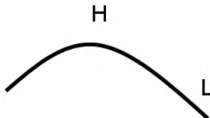
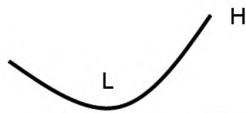


F0		
words	good MORning.	in LONdon ?
syllable structure		
annotation	H* L%	L* H%

Figure 4: Schematic representation of an utterance with a high accent followed by a low boundary tone (left) and a low accent followed by a high boundary tone (right). The first line indicates the F0 movement. The second line provides the text, with capital letters signaling the accented syllable. The third line is a representation of the syllable structure: The grey-shaded areas indicate the lexically stressed syllable. The fourth line gives the symbolic representation of the tones in AM-notation.

Both pitch accents and edge tones can be affected by two kinds of modifications: If a high tone (H) is considerably lower than a preceding high tone (without being a real low tone), it is said to be *downstepped*.¹⁴ If it is considerably higher in relation to the preceding high tone, it is said to be *upstepped*. Downstep is marked in transcriptions by inserting a “!” before the affected tone; upstep is indicated by a “^”.

¹⁴ Sequences of downstepped accents occur often in lists, as in:



“I bought apples, bananas, and melons.”

H* !H* !H* L-%

1.2 Linguistic markings of information-structural categories

Some models (including the model of German intonation adopted here) posit the existence of a third type of tone – the *phrase accent*. It is marked with a “-”, and always monotonal (i.e., H- or L-). Two functions have been attributed to the phrase accent: On the one hand it is said to control the pitch contour between the last pitch accent in the phrase and the boundary tone at the end of the phrase (see (16) below). On the other hand, it has also been assumed to be a demarcation marker of a particular prosodic unit: the intermediate phrase. Intermediate phrases (ips) are part of the so-called prosodic hierarchy, that is, the hierarchical organization of utterances into different-sized units, such as the grouping of syllables into words and of words into phrases. In this hierarchy, intermediate phrases are seen as the building blocks that make up the major intonational unit, the intonational phrases (IPs), which are in turn the units of which utterances are made (see Shattuck-Hufnagel & Turk, 1996 for an overview of different proposals for prosodic structure). The organization of the top end of the prosodic hierarchy is depicted in Figure 5. Where a phrase accent and a following boundary tone have the same pitch level (i.e., both H or both L), the German transcription system GToBI (German TOnes and Break Indices, Grice & Baumann, 2002) uses a simplified notation with only one tone (e.g., H-% instead of H-H%).

1.2 Linguistic markings of information-structural categories

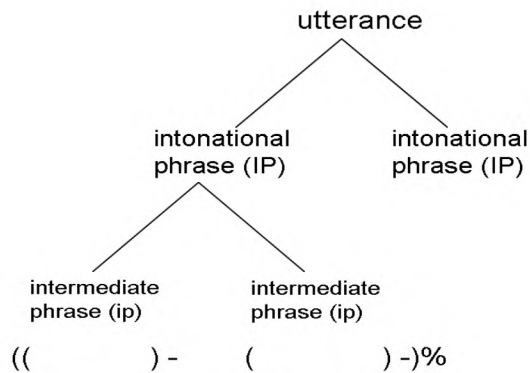
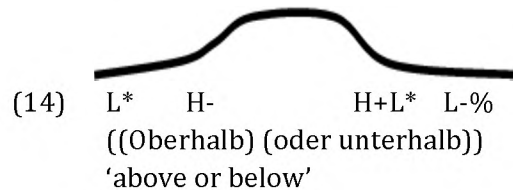


Figure 5: The upper end of the prosodic hierarchy. Edges of intermediate phrases are notated with a “-” sign and edges of intonational phrases with a “%” sign. Note that the branching is not necessarily binary.

An example of a phrase accent marking the edge of a (here high-ending) intermediate phrase is provided in (14).



(taken from the GToBI training materials)

One important assumption that follows from postulating an abstract level of tonal categories as in the AM framework is that there need not be a one-to-one correspondence between these tonal elements (i.e., pitch accents and edge tones) and the syllables or segments. The domain on which the tonal elements are realized can consist of only one syllable or of several syllables, as shown in examples (15) and (16) below, taken from Ladd (2008, p. 45f.).¹⁵

¹⁵ The contour’s annotation as L*+H L-H% was added by me.

1.2 Linguistic markings of information-structural categories

(15) A. I hear Sue's taking a course to become a driving instructor.



L*+H L-H%

B. Sue!?

(16) A. I hear Sue's taking a course to become a driving instructor.



L*+H L- H%

B. A driving instructor!?

Most of the examples presented so far contained only a single pitch accent, but of course longer utterances are likely to contain more than one accent, as in example (14) above. Of these, one accent is usually more salient than the others, and it is typically also the last accent in the phrase. This accent is referred to as the *nuclear* accent, and the accents that precede it are called *pre-nuclear* accents. Because of its prominence, the nuclear accent plays an important role in signaling pragmatic functions of constituents.

As already mentioned, the introduction of the standardized transcription system ToBI (Beckman & Ayers, 1997; Silverman et al., 1992) facilitated the description of intonation in an AM account for many researchers. Similar labeling standards for other languages were readily developed (see Jun, 2005)^{16,17}, among which the German version GToBI, which is also the annotation scheme used in this thesis. A complete ToBI transcription of a stretch of speech consists of at least three tiers: an orthographic transcription of the

¹⁶ Ladd (2008, p. 110ff.) critically discusses the (mis-)use of the ToBI system as an 'IPA style' annotation system for intonation and points out that only language-specific phonological abstractions should be mapped onto the speech signal.

¹⁷ In view of the many ToBI-like transcriptions for other languages, the annotation scheme for American English is now commonly referred to as "MAE-ToBI" (Mainstream American English ToBI).

1.2 Linguistic markings of information-structural categories

text, an annotation of the tones (pitch accents, phrase accents and boundary tones), and a tier with so-called break indices, which mark the perceived strength of the boundaries between different phrases in the utterance. However, in the remainder of this dissertation I will use the term “GToBI transcription” or “GToBI annotation” to refer only to the tonal elements.

The assignment of a label to a given word or constituent is based on both perceptual and acoustic criteria. For instance, the accent L*+H is described as: “a low target within the accented syllable is followed by a rise, starting late in the accented syllable and reaching its peak on the next syllable (or sometimes later). [T]he perceived pitch of the accented syllable is low” (Grice, Baumann, & Benz Müller, 2005). Table 1 below gives schematic representations of the six basic¹⁸ pitch accent types according to GToBI, along with a description of their respective characteristic features as stated in Grice et al. (2005).

¹⁸ As noted before, H* accents may be modified phonologically by up- or downstep.

1.2 Linguistic markings of information-structural categories

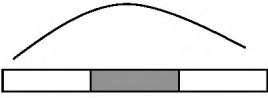

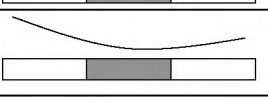
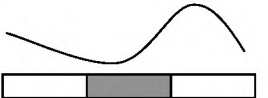


Symbol	Schematic representation	Characteristic features
H*		<ul style="list-style-type: none"> - 'peak accent' - perceived as relatively high - may be preceded by shallow rise
L+H*		<ul style="list-style-type: none"> - 'rise from low up to peak accent' - peak preceded by low pitch target, sharp rise in accented syllable - peak often late in accented syllable
L*		<ul style="list-style-type: none"> - 'low accent' - accented syllable is local pitch minimum low in speaker's range
L*+H		<ul style="list-style-type: none"> - 'valley accent plus rise' - low target within accented syllable followed by a rise, starting late in accented syllable - peak on next syllable (or later)
H+L*		<ul style="list-style-type: none"> - 'step-down from high to low accent' - accented syllable low at or very near bottom of speaker's range - preceded by high pitch target on preceding syllable
H+!H*		<ul style="list-style-type: none"> - 'step-down from high to mid accent' - higher pitch on pre-accented syllable - accented syllable around the middle of the range - often continuous fall from pre-accented syllable through acc. syllable to final syllable in phrase

Table 1: Schematic representation of the six basic GToBI pitch accent types. The grey-shaded area indicates the lexically stressed syllable; the white areas preceding and following it represent the pre- and post-accentual syllable, respectively.

Despite these seemingly clear criteria, even trained labelers still experience difficulties in agreeing on the same labels for a given accent. This is shown by intertranscriber consistency studies that have been conducted both with the

1.2 Linguistic markings of information-structural categories

original English ToBI system (Pitrelli, Beckman, & Hirschberg, 1994; see also Syrdal & McGory, 2000) and the German GToBI system (Grice, Reyelt, Benz Müller, Mayer, & Batliner, 1996). While the agreement on the presence or absence of a pitch accent was reasonable (81% in the English study, 87% in the German study), agreement on the type of pitch accent was much lower (64% in the ToBI experiment and only 51% in the GToBI experiment). Considerable inconsistency is also reported by Braun (2006), who asked trained labelers to annotate topical referents in German sentences choosing only among the three accents H*, L+H*, and L*+H.

As for edge tone labels (i.e., phrase accents and boundary tones), the results are somewhat more encouraging: In Grice et al.'s (1996) study, labelers agreed on both the location and type of label in 86% of the cases. In Pitrelli et al.'s (1994) study on English, there was an agreement in about 90% of the cases regarding the presence or absence of a phrase accent or boundary tone, and – when they agreed on the presence of an edge tone – labelers used the same label in 73% (for phrase accents) and 79% (boundary tones) of the time, respectively.

Still, especially the disagreement on pitch accent type labels may cause some people to question the usability of the (G)ToBI system, and may fuel the criticism that phonological analyses are often too impressionistic and without sound empirical basis. Yet the mere fact that people have difficulties using such a system does not mean that it is not desirable and even necessary to have an analysis of the intonational phonology of a language. Just as much as knowledge of the classes of sounds of a language (its phoneme inventory) is needed to adequately describe it, it is necessary to understand what the tonal categories are, in order to be able to understand how these are used to express certain meanings, such as marking a constituent as the topic of a sentence. And, as Ladd (2008) notes, it should be recognized that phonological categories could in principle be instrumentally validated (p. 12). In the case of GToBI accent categories, there should, for example, be a measurable F0 minimum within the accented syllable of a L*+H accent; or there should be a clear pitch fall from a high target into the accented syllable of a H+L* accent. Of course, labelers use these criteria implicitly in their judgments, but we may

1.2 Linguistic markings of information-structural categories

also measure the parameters in question directly to have more objective evidence for the correct application of these criteria. Furthermore, acoustic data of this kind allows for the comparison of realizations of a given accent category across different groups of speakers, such as native speakers and second language learners (as has been done by Mennen, 2004) – or speakers of different ages, as in this dissertation.

However, it needs to be acknowledged that manual labeling (i.e., deciding on the type of label for a given accent or edge tone) is very time-consuming. This often has the consequence that researchers working in this type of framework need to accept smaller sample sizes than are common in most psycholinguistic experiments, for instance. This is even more the case when dealing with natural data as opposed to scripted speech. When, in addition to assigning phonological labels, acoustic landmarks like turning points have to be annotated as well, this process may take even longer. A complicating factor is that the determination of these landmarks is not always without problems. For example, certain segments (e.g., fricatives such as /v/) cause so-called microprosodic variation in the pitch curve, making it sometimes difficult to identify the location of turning points (such as the F₀ minimum in L*+H accents) unambiguously. Here labelers need to find ways to abstract away from these phenomena. On the other hand, the current practice of having multiple annotators code the same data set and compare and discuss their labels may not be any less labor-intensive.

The preceding discussion points to some of the methodological difficulties involved in the research of intonation in speech production. Additional problems arise from the fact that for any given language, there is typically not one agreed-upon inventory of tonal elements. For German, several proposals have been made (for an overview, see e.g., Grice & Baumann, 2002; Kügler, 2007), of which GToBI is only one. Often there is dispute as to whether a particular accent constitutes a separate phonological category, or whether it is merely a phonetic variant of another accent. (The (G)ToBI category L+H* is a case in point: While some claim that it is an accent in its own right, others assume that it is actually a variant of H*.) A discussion of all the issues involved in these debates is beyond the scope of this thesis; the point to be made here is that

1.2 Linguistic markings of information-structural categories

whenever one decides to work with one system and not the other, there is not only the danger of inaccurate or subjective annotation according to the rules of that system, but also the possibility that one is analyzing the data using the “wrong” system of categories. An additional but related problem confronts the researcher dealing with first language acquisition: To what extent is it appropriate to apply adult categories to child speech? In view of these problems, it is even more beneficial to gather quantifiable acoustic data. Such data allows us to also evaluate our findings from a perspective that is independent of the phonological category system that we have chosen to use.

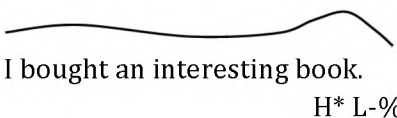
This section has provided a general introduction to intonation and to the AM framework and the GToBI annotation in particular. A number of methodological issues were raised, some of which will be addressed in the subsequent chapters. The following section now discusses some of the (purported) mappings between information-structural categories and specific pitch accent types.

1.2.2.2 The role of pitch accent types in signaling information structure

Picking up on the discussion of linguistic markings of information structure above, we can now specify the different pitch accent types more accurately in terms of AM notation. I will first discuss the marking of information status of discourse referents, and then turn to the marking of the pragmatic functions of focus and topic. We will also briefly touch upon the role of phrase accents and boundary tones in discourse. This section is mainly intended as a primer on theoretical proposals, but includes references to relevant empirical work as well.

The accent on new referents is typically taken to be a high pitch accent like H* (e.g., Baumann & Grice, 2006; Gussenhoven, 1984; Pierrehumbert & Hirschberg, 1990), illustrated in (17) (repeated from (1) above).

1.2 Linguistic markings of information-structural categories

(17) 
I bought an interesting book.
H* L-%

As explained before, given referents are said to be realized without an accent, or to be “deaccented” (Ladd, 1980). The term “deaccentuation” reflects the idea that there is no accent where one would otherwise expect one (i.e., if the referent were not given).¹⁹ In most cases, the accent shifts to the left compared to the non-deaccented position. Compare the two exchanges in (18) (taken from Ladd, 2008, p. 270).

- (18) a. A. Why didn’t you read the article I gave you?
B. I can’t read GERMAN.
- b. A. The only article on this is in German.
B. I can’t READ German.

Note that the accent on *read* in (18b) is not an expression of this constituent being focused; there is no context question about the verb *read*, nor is it contrasted with, say, *write*. The prominence is a mere consequence of the deaccenting of *German*.

Although deaccentuation is seen as the canonical realization for given referents, alternative proposals for certain pitch accent types have also been put forward. Pierrehumbert and Hirschberg (1990) assume that information that the speaker does not intend to add to the speaker’s and hearer’s mutual beliefs (i.e., given referents) are realized with a L* accent. In one of their examples (their 18), the speaker is asked to give a list of things he wants for his

¹⁹ In the remainder of this dissertation, I will use the term “deaccented” rather than “unaccented”. Note that this decision does not entail any theoretic position; it is merely used for practical reasons, since there is for instance no corresponding noun to “unaccented” as there is for “deaccented” (deaccentuation).

1.2 Linguistic markings of information-structural categories

birthday. Having already having indicated his desire for a Pavoni espresso machine at some earlier point, he utters (19).

(19) Well, I'd like a Pavoni...
L* L* L* L-H%

The L* accent features also in earlier theories, for instance by Brazil (1975): He claims that if a speaker says something that he wants the hearer to take as already established in the discourse (his “referring”), this is indicated with a rise (probably L*+H). An alternative is a fall-rise (H* L-H%), the same accent that is also suggested by Gussenhoven (1984) as the appropriate tone for “selecting” a variable from the discourse, which seems comparable to Brazil’s referring. Thus most accounts seem to converge on the assumption that if given referents carry a pitch accent, they are realized with a low pitch accent or at least some rising pitch trend.

Some corroborative empirical evidence for the correlation between information status and intonation comes from online perception experiments using the visual world paradigm (e.g., Dahan, Tanenhaus, & Chambers, 2002). In this paradigm, participants are instructed to move objects (referents) on a computer screen while their eye fixations are monitored. Some of the referents carry names whose onsets are initially phonemically ambiguous (e.g., candy and candle). What is manipulated is the intonational realization of those words, and their information status in discourse (e.g., previously mentioned/not mentioned). Participants’ eye fixations reveal their interpretation of the auditory (intonational) information they hear. Using this set-up, Chen, Den Os and De Ruyter (2007) observed for English that participants tend to fixate the referent that was new more when it carried a falling accent (H* L- or L*+H L-) than when it was deaccented or was realized with a rising accent (L*+H). Conversely, when the referent was given, L*+H accents and deaccented realizations triggered more looks to it.

Intonational realizations of referents with an intermediate degree of givenness have been discussed to a much lesser extent. Pierrehumbert and Hirschberg’s (1990) theory of intonational meaning mentions that down-

1.2 Linguistic markings of information-structural categories

stepped accents (!H* and H+!H*) signal that a referent is inferable (i.e., accessible via inference) from the mutual beliefs of the speaker and hearer. For German, Baumann and Grice (2006) have recently put forward a proposal that the accessible status of a referent can be signaled with the accent H+L*. This hypothesis is based on the results of a corpus study (Baumann, 2006) and an offline perception experiment (Baumann & Grice, 2006). The corpus consisted of 22 short texts (250 sentences altogether), which were taken from the economics section of the German newspaper *Frankfurter Rundschau* and which were read by a single male speaker of standard German. Baumann analyzed the information status (new, given, or accessible) and the type of pitch accent of the last referring expression in each sentence, provided this last referring expression was a full noun phrase. An example is given in (20).

- (20) Die vor einem Jahr eröffnete Fabrik geht gerade
the before one year opened factory goes presently

zum ZWELschicht-Betrieb über.
to the TWO-shift operating system over
status: New
accent: H*

'The factory that opened one year ago is presently changing to two-shift operation.'

Only low-ending utterances (i.e., those ending with a L-% boundary tone) were looked at. The analysis of 106 items²⁰ overall showed that more than

²⁰ The reduction to 106 items (as opposed to the expected 250) is due to two factors: First, all utterances that ended in a high boundary tone were excluded (i.e., questions). Second, Baumann excluded items in which the last referring expression was deaccented but followed by a nuclear accent on another constituent, as in *Die New Yorker Aktienbörse hat am Freitag ihre Rekordfahrt FORTgesetzt* ('On Friday the New York

1.2 Linguistic markings of information-structural categories

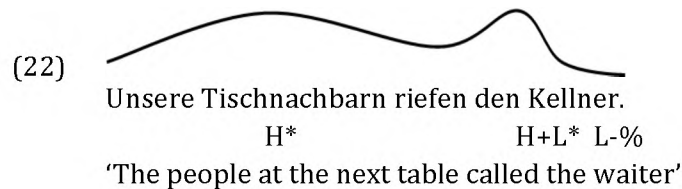
55% (34 out of 61) of the accessible referents were realized with a H+L* accent, suggesting that it is a suitable marker for accessible referents. In contrast, the accent H*, which is typically assumed to signal newness, was used in only 2 out of 61 cases. This suggests that H* is not a prototypical accent for accessible referents. Further evidence for the role of H+L* as a marker of accessibility was presented in an offline perception experiment, in which participants had to rate the appropriateness of different intonation contours in varying contexts on a scale from 1 to 7. In the experiment, subjects read and simultaneously listened to 40 short paragraphs in which the target referents were embedded. The contexts were assumed to render the target referents accessible, for instance because they had been mentioned several clauses before (textually accessible) or because they were for instance part of a larger scenario, as in (21) below. Here, the target referent is *the waiter*, which is embedded in a restaurant scenario.

- (21) The restaurant was excellent. It was already a pleasure to read the menu.
Nonetheless, we couldn't have ordered everything we would have liked. The people at the next table called the waiter. They had already drunk two bottles of champagne.

The target referents were realized with three different intonation contours: with a H*, a H+L* or deaccented. The contour containing H+L* is shown in (22).

stock exchange continued its record-breaking run'), where *Rekordfahrt* is the last referring expression in the sentence, but the nuclear accent occurs on the predicate *fortgesetzt*.

1.2 Linguistic markings of information-structural categories



The listeners' task was to judge the contextual appropriateness of the target sentence's intonation patterns. Baumann and Grice found that deaccentuation was in most cases the preferred realization. However, the H+L* accent was considered equally acceptable for textually accessible (previously mentioned) referents, and it was always judged to be a better realization than H* in all contexts. On the basis of these findings, Baumann and Grice (2006) suggest the following mapping of information status and accentuation type for German:

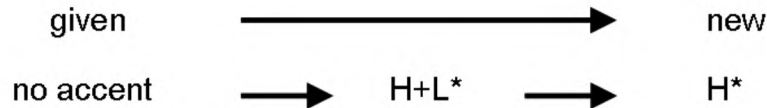


Figure 6: Proposed mapping between information status and accentuation type in German (adapted from Baumann & Grice 2006, p. 1655).

These results show that listeners are sensitive to a referent's information status in discourse, and that they have intuitions about its appropriate intonational realization. Together with findings from online experiments (Chen et al., 2007), this seems to indicate that the choice of pitch accent type for a discourse referent is not arbitrary, but pragmatically governed.

Turning now to the marking of pragmatic function, we find that the commonly observed falling movement on focal referents translates in many accounts (e.g., Steedman, 2000; Uhmann, 1991), into a high pitch accent (H*), typically followed by a low phrase accent (L-):

1.2 Linguistic markings of information-structural categories

(23) A: Which result did Marcel prove?



B: Marcel proved [COMPLETENESS]_F.

H* L-%

(Steedman, 2000, p. 657)


It has also been suggested that speakers tend to use L+H* rather than H* accents if they wish to convey some kind of contrast with potential alternatives (Bartels & Kingston, 1994; Pierrehumbert & Hirschberg, 1990). Results from eye-tracking experiments lend some support for this hypothesis, as it has been found that listeners have a strong tendency to interpret L+H* accents as contrastive (Braun, Weber, & Crocker, 2004; Ito & Speer, 2008; Watson, Tanenhaus, & Gunlogson, 2008). Studies on focus marking in German furthermore suggest that the size of the focus domain can also modify the realization of a pitch accent. Baumann, Becker, Grice and Mücke (2007) observed that speakers produced H* accents on focal referents more often with downstep (!H*) in broad-focus contexts (i.e., in sentences that were uttered as a response to a question like *What's new?*), whereas they produced more upstepped H* accents (^H*) in contrastive contexts (e.g., *Does Marlene want to peel a potato? – Marlene wants to peel [a banana]_F*). Similar results are reported by Féry and Kügler (2008).

Unlike focused referents, topical referents are assumed to be produced with a prenuclear accent typically followed by a high phrase accent. This prenuclear accent has been described differently as either L+H* (e.g., Gundel & Fretheim, 2004; Steedman, 2000; Vallduví & Engdahl, 1996) or L*+H (e.g., Jacobs, 2001; Steedman, 2000), but a simple H* has also been suggested (Büring, 2003, 2007). For Steedman (2000, 2007), the choice between L+H* and L*+H accents hinges on whether the topic is mutually “agreed” or “non-agreed”, where a non-agreed topic is one that the speaker assumes to be contentious. This distinction seems to some extent related to the distinction of contrastive vs. non-contrastive, if one assumes that a non-agreed topic is likely to contrast with some other alternative. For German, it has been claimed that in particular contrastive topics are often realized with the so-called *hat pat-*

1.2 Linguistic markings of information-structural categories

tern. This pattern, which was originally described by Cohen and t'Hart (1967) for Dutch, is characterized by an initial rise (on the topic constituent), a sustained high pitch and a final fall onto the nuclear accent, resulting in a pitch curve that bears resemblance with a side-view on a hat. If we translate the answer in example (12) into German, (24) is a likely realization.

(24) A. What about the rest of the class? Have you heard anything?


B. [JILL]_T ist jetzt LEHRERIN.
L*+H H* L-%
'Jill is now a teacher.'

An empirical investigation of the production and perception of contrastive and non-contrastive topics was conducted by Braun (2006). In her production experiment, participants read target sentences embedded in short paragraphs, which provided contrastive or non-contrastive contexts for the target sentences. The same sentence was produced in both context types. This design allowed for a direct comparison of the two conditions for each subject individually. Braun found that all topics were realized with the rising accents L+H* and L*+H. However, the two accent types were equally distributed across both conditions, and there was also no difference in the frequency of hat patterns between the two conditions. Contrastive and non-contrastive topics were thus not differentiated phonologically. On the phonetic level, however, subjects distinguished contrastive topics from non-contrastive ones by using higher peaks, a greater pitch excursion and later alignment of the peaks. Hence Braun's results do not support the hypothesis that contrast is signaled by a particular pitch accent type, as suggested by Steedman (2000, 2007), but rather by gradient phonetic modification of an accent.

Finally, I shall briefly say something about phrase accents and boundary tones and the discourse-structuring role that has been ascribed to them, although edge tones are not the primary concern of the studies presented in this dissertation. In an analysis of English intonation, Brown, Currie and Kenwor-

1.2 Linguistic markings of information-structural categories

thy (1980) distinguish low and not-low terminals, that is “low-ending” utterances and “not low-ending” utterances (which includes high ending ones). They assume that low ending utterances occur at the end of a turn and indicate that the speaker does not have any more to say on a topic. Not-low ending utterances, on the other hand, are commonly used to indicate that the speaker is about to say more (p. 30). In a similar vein, Pierrehumbert and Hirschberg (1990) assume that H phrase accents (H-) indicate that the current intermediate phrase should be taken to be “part of a larger composite interpretative unit” (p. 302). That is, it should be interpreted together with the following ip. In contrast, L phrase accents emphasize that the current phrase is separate from the following one. Accordingly, the authors suggest for boundary tones that a high boundary tone (H%) conveys that an intonational phrase is “forward-looking”: the speaker wants the hearer to interpret the utterance with attention to subsequent utterances. A low boundary tone (L%), on the other hand, is not specified for directionality in the same way.

As explained before, Baumann and colleagues looked in their analysis of information status marking only at utterances ending in low boundary tones; it would be interesting to see how the speakers’ use of pitch accents and boundary tones may interact in ongoing discourse.

Before we turn to a review of information-structural markings in child language, however, note that in the studies so far, all findings have been obtained either from read-out texts, or from perception experiments in which participants listened to read-out stimuli, which were presented in carefully created but rather restricted contexts. This does not render the findings invalid, but leaves the evidence for a direct mapping between certain information-structural categories and particular accentuation types indirect. To what extent these relations hold for natural speech production, is still an open question.²¹ Yet most of our everyday speech communication is unscripted: People

²¹ For English, Calhoun (2006) has looked at the relationship between information structure and prosody in spontaneous speech and concludes that information structure imposes only a *probabilistic* – rather than a deterministic – constraint on the way words are mapped with prosodic structure.

1.3 The acquisition of linguistic markings of information structure

speak without being given a text to produce in advance; they have to construct everything online. This has a number of effects we do not typically see (or rather hear) when people read out aloud. Speakers produce pauses and hesitations, and they may finish off their sentence different from how they started it, often resulting in ungrammatical constructions. Spontaneous speech is also notorious for its phonetic reductions: A phrase like *because if* may be pronounced [k^hzɪf] (Johnson, 2004, p. 31). Seeing how everyday speech deviates from “the norm” in so many ways, we may expect some differences between scripted and unscripted speech also in intonation. This question will be investigated in chapter 6. It is therefore important to investigate the relationship between information structure and intonation also in more natural settings – not least because children, too, engage in and experience mainly unscripted communication, and will adopt this as their target model.

The previous sections have provided some background on information structure and linguistic markings of information structure, in which way different pitch accent types have been suggested to mark the information status and the pragmatic function of discourse referents. The next section focuses on the *acquisition* of these linguistic markings, and reviews previous findings on what is known about children’s use of language to “co-construct a universe of discourse” (Hickmann, 2003, p. 41).

1.3 *The acquisition of linguistic markings of information structure*

The discussions in sections 1.1 and 1.2 have shown that choosing the appropriate form to encode a referent involves at least three skills. First, the speaker needs to be able to make an assessment of the addressee’s perspective. This involves knowing whether the addressee can identify the referent, and estimating the addressee’s awareness of that referent at that particular moment in the discourse. Second, the speaker needs to have learned the necessary linguistic forms, that is, the various referring expressions, syntactic forms and intonation patterns that are used to mark the information status and the pragmatic function of a referent. Third, the speaker has to know when and how to use these forms. These three requirements make the learning of how

1.3 The acquisition of linguistic markings of information structure

to mark information structure a complex task for a child. This section provides a brief review of the most important findings related to these three aspects. The first part reports research on children's understanding of others' perspective (section 1.3.1). The second part summarizes what is known about children's knowledge of morphosyntactic forms and how they make use of them to mark information structure (section 1.3.2). The last part of this section is devoted to the few studies on intonation (section 1.3.3). It also begins with an overview of the available evidence regarding children's command of the necessary forms (i.e., intonation contours), and goes on by looking at how children use these forms to signal information structure.

1.3.1 Assessment of others' perspective

Recent research provides evidence that even very young children understand what another person may perceive or know, and demonstrate this understanding also in their communication. One important prerequisite to understanding that someone else can have different knowledge about the world is the ability to take a perspective different from one's own. In the domain of visual perspective taking, recent evidence shows that children as young as two years of age can take a non-egocentric view. For example, Moll and Tomasello (2006) found that 24-month-olds (but not 18-month-olds) knew that when an adult asked for the children's help to look for a specific object (*Where's the toy?*), that he was not looking for the object that was visible to both the child and the adult, but for an object that only the child could see. The children demonstrated that they had understood that they saw an object the other person could not see (so-called Level-1 perspective taking, cf. Flavell, 1974; Flavell, Everett, Croft, & Flavell, 1981).²² Children of two years of age have been shown to know about the perceptual availability of discourse referents – which is a prerequisite for evaluating and marking its information status. They also keep track of what their interlocutor has experienced in the immediate past. Tomasello and Haberl (2003) found that when an adult asked excitedly

²² Level-2 perspective taking refers to the child's ability to understand that she and another person can see the same object simultaneously from different perspectives.

1.3 The acquisition of linguistic markings of information structure

to be handed a toy from an array of three without indicating which one (*Oh look! Look at that one! Can you give it to me?*), 1 and 1.5-year-old children reliably handed the toy that was new to the experimenter, but equally old for them (see also Moll, Koring, Carpenter, & Tomasello, 2006; Moll & Tomasello, 2007). Thus, very young children have awareness of which objects another person is and is not familiar with, and they seem to take the interlocutor's knowledge into account when they engage in interaction.

These abilities have also been found to have repercussions on early language use. For instance, Wittek and Tomasello (2005) showed children of 2.0, 2.5 and 3.5 years a clown puppet sweeping the floor with a broom, and putting the broom away (on a shelf that was 1.5 m away from the testing location). The "oblivious" experimenter then entered the scene and uttered *It's pretty dirty in here. I bet the clown wants to sweep the floor.* This was followed by a question that either indicated knowledge about the object (*What happened to the broom?*) or not (*What do we have to get?*). A third condition was a so-called contrast question (*Do we need a vacuum cleaner?*). When asked a specific question about the object, the older children (2.5 and 3.5 years) tended to use null forms or pronouns to refer to that object (e.g., *on the shelf* or *it's on the shelf*). In contrast, when they were asked more general questions they tended to use lexical nouns (e.g., *a broom* or *no, a broom*). The children showed sensitivity to the interlocutor's knowledge state, as indicated by the experimenter's question. It was also observed that overall pointing to the location of the object occurred much more frequently with the more presupposing forms, that is, with null forms and pronouns (these forms presuppose more awareness on behalf of the listener). This behavior indicates that the children had some awareness that supplementary (extra-linguistic) information is often necessary with these forms. However, it also shows that they were still inclined to make use of real-world context, even though the extra-linguistic referent was not directly perceptually available to their interlocutor.

1.3 The acquisition of linguistic markings of information structure

1.3.2 Morphosyntax

With respect to the repertoire of linguistic forms, morphosyntax seems to be in place fairly early: The usage of both definite and indefinite articles as well as pronouns has been documented from about 2;0 years (e.g., Brown, 1973; Keenan & Klein, 1975). The majority of these forms seem to occur initially in naming constructions in connection with paralinguistic markers such as pointing or gaze (e.g., *That's a bunny!*). There is furthermore naturalistic and experimental evidence that indefinite articles are used as markers of non-specificity from early on (e.g., *I want an apple*) (Brown, 1973; Karmiloff-Smith, 1979; Maratsos, 1974, 1976). However, when it comes to using these forms appropriately to mark information-structural dimensions, in particular the *information status* of referents, the evidence is less clear. Generally, analyses of natural conversations and studies in which children were asked to tell narratives of personal experience or conventionalized stories such as fairy tales (e.g., Bennett-Kastor, 1983; Keenan & Klein, 1975) find earlier correct uses of referring expressions for information status marking than do controlled experimental studies. The problem with the former type of studies is that they normally do not allow control over the content being talked about. Especially in conversational settings, participants talk mostly about the here-and-now, making it difficult to disentangle deictic (exophoric) uses of determiners and pronouns from discourse-internal (endophoric) ones. In order to assess children's ability to mark information status endophorically, it is necessary to set up a situation in which they have privileged knowledge that they communicate to a listener. Studies that were designed this way have typically elicited narrations using picture sequences that are only visible for the child. The majority of these studies report that correct endophoric uses of referring expressions emerge only rather late, around the age of seven (e.g., Hickmann, Hendriks, Roland, & Liang, 1996; Warden, 1976). One study that explicitly manipulated mutual knowledge was done by Kail and Hickmann (1992), who directly compared narratives from six-, seven- and nine-year-old French children in two conditions: one in which the children were looking at a picture book together with their interlocutor (*mutual knowledge*), and one where the

1.3 The acquisition of linguistic markings of information structure

children were looking at the picture book by themselves with the interlocutor blindfolded (*no mutual knowledge*). While children of all ages were found to differentiate between the two situations – as evidenced by the fact that the distribution of definite and indefinite forms was different for the two conditions – a clear developmental trend was observed. In the no mutual knowledge condition, the youngest age group (six years) did not show a preference for indefinite forms for referent introduction over the other forms. It was only with 9 years that children’s first mentions in that condition were appropriate.²³ Similar results were found by Hickmann, Hendriks, Roland and Liang (1996), who analyzed introductions of new referents in narratives by four-, seven- and ten-year-old children in a cross-linguistic study. For German, they found that pre-school children (four to five years old) introduced new referents inappropriately (e.g., with definite or bare NPs) in 45% of the cases. There was a significant decrease to the next age group (seven-year-olds), in which errors occurred only in a quarter of the cases (26%), but even ten-year-olds still produced definite determiners for first mentions 15% of the time.

These results seem to suggest that children make “egocentric” errors, that is, they apparently fail to understand that their interlocutor may have a different perspective (i.e., may not be able to identify a referent). Yet this hypothesis is in conflict with the findings discussed in the first part of this section, which show that children much younger are sensitive to the interlocutor’s point of view. It seems therefore unlikely that five-year-olds would fail to take the addressee’s perspective into account. One possible explanation for these errors has been put forward by researchers who approach language acquisition from a functional perspective (e.g., Karmiloff-Smith, 1979, 1981). They suggest that children who make egocentric errors have not yet learned that determiners have – among others – the function to mark information status intralinguistically. Instead, it is argued that children use definite articles and pronouns de-

²³ An interesting finding was that eleven-year-olds tended to use indefinite introductory forms also in the mutual knowledge condition. It appears that by that age children have learned the conventions of narration, which may be said to be more a cultural skill than linguistic ability (see also Kail & Hickmann, 1992, p. 90).

1.3 The acquisition of linguistic markings of information structure

ictically, anchored to the referent in the real world. Only later do they become aware of the information status marking function of determiners. However, the “final verdict” on why even older children make these kinds of errors has not been reached.

Children’s sensitivity to discourse referents’ pragmatic functions as topic or focus of a sentence has been mainly looked at from a syntactic perspective. Most studies find evidence for early emergence of topic and focus marking. For instance, in a case study, Scollon (1979) observed that an English acquiring child at the one-word stage would often produce consecutive one-word utterances in which the first one designated the topic, and the second one constituted the predication made about the topic (e.g., *Finger. Touch*), in a situation where the child was about to touch the microphone with her finger. Similar constructions in the two-word stage (around two years) are reported by D’Odorico and Carubbi (2003) for Italian children. De Cat (2003, 2009) shows that French children under three use left and right dislocation constructions (which are obligatory markers of topics in French) only in contexts where the dislocated element can be licensed as a topic.

In sum, research shows that from very early on, children are able to assess their interlocutor’s knowledge. Children as young as 1.5 years adapt their non-verbal behavior according to their understanding of what the interactant knows. When they use language, children from 2.5 years onwards take discourse availability (i.e., preceding linguistic context) into account and are more likely to use more informative linguistic forms (full noun phrases) when they assume that their listener is not familiar with a referent. They also use syntax to mark referents as the topics of sentences. However, the evidence seems to converge at a prolonged acquisition of the correct use of morphosyntactic forms to mark information status, in particular the marking of an unidentifiable referent with an indefinite form. The reason for this is not entirely clear. What can be concluded, however, is that development in this domain clearly continues into the school years. Adult-like marking does not seem in place before age seven.

1.3 The acquisition of linguistic markings of information structure

In the last part of this section, we now turn to children's use of *intonation* to mark information structure, which has been studied to an even lesser extent.

1.3.3 Intonation

In contrast to referring expressions, adult-like command over intonation appears to be a later development. Several studies report that rising contours in particular may pose some difficulties for children. Local (1980), for instance, found evidence for ongoing development of the intonation system through the school years. Analyzing spontaneous speech of children from northeast England (Tyneside), he found that during the course of their sixth year, the children began to produce fewer falls and more level and rising tones, which he interprets as an indication that the children are getting closer to the adult inventory of pitch accents. (In Tyneside English, levels and rises are particularly frequent, because they are used with declarative sentences.) However, Local does not provide any further information about the exact shape of these tones.

More acoustic detail on children's ability to produce different intonation contours can be found in studies from the field of speech and language pathology. Snow (1998; see also Snow, 2001) asked four-year-olds to imitate falling and rising contours modeled by an adult speaker. The sentences used for imitation were short utterances like *The cat has a BOTTLE* (falling contour) or *Did you take your BOTTLE?* (rising contour). He found that children had the tendency to substitute the modeled rising contour with a falling one. When they produced rising contours, they did so with longer word durations coupled with narrower pitch range. This means that their rate of pitch change (or slope) was slower than that of the adult model. In contrast to that, they did not have any problems with imitating falling contours. Similar findings are reported by Loeb and Allen (1993), who also used an imitation paradigm, and Patel and Grigos (2006), who elicited semi-spontaneous statements and questions in a game-like format. Patel and Grigos introduced four-, seven- and eleven-year-old children to puppets representing characters and objects, such as "Bob" (a character modeled after SpongeBob SquarePants) and "bot" (a

1.3 The acquisition of linguistic markings of information structure

robot). The children's task was to instruct the experimenter to perform a task involving one character and one object. A short contextual scenario served as a prompt to elicit a phrase as either a question or a statement. For example, in order to elicit the phrase *Show Bob a bot*, the children were told that Bob was lonely and needed something to play with. The experimenter (Experimenter A) would then ask the child *What should I show Bob?*. The elicitation procedure for the corresponding question (*Show Bob a bot?*) was more complicated: The children were told that Bob needed help with something, and one of the experimenters (Experimenter A) was offering a potential solution by saying *Show bob a bot*. The other experimenter (Experimenter B) would then look puzzled and ask the child *Does this make sense? Ask her what she wants to do?*, and the child would be expected to ask *Show bob a bot?*.

Patel and Grigos found that four-year-olds were not able to reliably signal these questions by raising F0 towards the end of the phrase. Instead, they increased the duration of the final syllable. Seven-year-olds were comparable to older children (eleven years) in their use of F0 to indicate the contrast between questions and statements, but still showed a great deal of variability (see also Wells, Peppé, & Goulondris, 2004 for similar results).

Assuming that speed of pitch change reflects physiological effort (cf. Ohala & Ewan, 1973), the findings just discussed suggest that for children under seven, final rising contours are more difficult than falling contours. Xu and Sun (2002) found that adults, too, take longer to increase pitch than to decrease it. If rising contours are more difficult for adult speakers, it seems reasonable that they are even more demanding for children whose laryngeal and respiratory systems are still maturing (Stathopoulos & Sapienza, 1997).

However, both the imitation studies (Loeb & Allen, 1993; Snow, 1998) and Patel and Grigos' (2006) experiment arguably raise some questions regarding their ecological validity. As for the former, it has yet to be shown to what extent imitated intonation contours are comparable to natural speech production in children.²⁴ As for the latter, Patel and Grigos (2006) argue that the ut-

²⁴ Snow (1998) also collected semi-spontaneous data from the same children in order to compare it with the imitations. However, he was only able to find 19 cases in which

1.3 The acquisition of linguistic markings of information structure

terances they elicited “may better approximate the child’s prosodic control abilities compared to directly imitated productions” (p. 1312), but the task evokes utterances which can be expected to still differ from spontaneous productions. The authors acknowledge that especially the four-year-old children needed additional cues and prompts, and in many cases, the expected response had to be modeled by the experimenter. This is maybe not surprising for two reasons. First, the child had to produce an echo question on behalf of the experimenter (Experimenter B), who could have asked the question herself. This procedure may not be very intuitive for a child. Second, the authors do not cite any evidence that suggests four-year-olds do normally use echo questions. If this is not the case, it may have been difficult for them to produce the adequate contour.

Despite these limitations, the available evidence suggests that even school children may not have fully acquired all forms (i.e., pitch accents and intonation contours) that adults use in information structure marking. For instance, if rising contours are indeed more difficult for children, this may have consequences for the way they mark topical referents – which are typically associated with rising accents. This brings us to the last aspect of this section: How do children use intonation to mark information structure?

Anecdotal evidence from early corpus studies points to appropriate accentuation of new referents and deaccentuation of given referents. Wieman (1976) reports that in two-word utterances children (aged between 1;9 and 2;5) deaccented previously mentioned referents, as in (25).

the children had produced any of the target words from the imitation routine (e.g., *home, bottle*) in sentence-final position. Of these 19 cases, 18 were falling contours. The falling contours were comparable with the imitated contours in terms of pitch range and duration. But since there were no rising contours in the spontaneous speech data, it is still unclear for these contours if the differences that were found between adult models and children in the imitation task generalize to spontaneous speech.

1.3 The acquisition of linguistic markings of information structure

- (25) Mother: What's in the street?
Child: FIRETRUCK street.

Wieman concludes from these observations that “like adults, children operate with an appreciation of what is new in their utterances and apply stress accordingly” (Wieman, 1976, p. 286). Halliday (1983) documents similar cases in his diary study of one child. These natural data hint at some early use of intonation as a marker of information status, but their generalizability is limited. However, there have been some experimental studies as well: Hornby and Hass (1970) tested how four-year-old children described pairs of pictures which differed from each other in one element. The second of the two pictures contrasted with the first one in that it showed either a different agent (e.g., a boy riding a bike vs. a girl riding a bike), a different action (e.g., a man washing a car vs. a man driving a car), or a different object (e.g., a girl petting a cat vs. a girl petting a dog). The contrasting elements were found to receive an accent, showing that children recognize contrast in consecutive events and mark contrastive elements prosodically. However, the study leaves unanswered the question of *how* exactly contrastive elements were realized differently from non-contrastive ones. The coding procedure is described with “sentences were scored for contrastive stress” (Hornby & Hass, 1970, p. 397), which is probably meant to refer to a more pronounced form of accent, yet this is not explicitly defined. The study is illustrative of the developmental studies of that time, in which the terms “stress” or “accent” were often not clearly defined, and the occurrence of these suprasegmental events was claimed exclusively on the basis of auditory impression (see also MacWhinney & Bates, 1978). While this is in itself not necessarily problematic – it has been common practice in the tradition of the so called British school of intonation (see Cruttenden, 1986, for an overview of this research tradition) – assessment of intonation purely by ear needs to follow clearly specified criteria, and usually requires a substantial amount of experience with acoustic data. These two conditions often appear not to be met in these studies, and at least the criteria

1.3 The acquisition of linguistic markings of information structure

that were used are not reported. With the advent of digital recording and generally available speech analysis software, more recent studies tend to include acoustic measurements of pitch, duration and sometimes intensity. I will first review two studies that analyzed child intonation in purely acoustic terms, and subsequently discuss a phonologically-oriented study.

Müller, Höhle, Schmitz and Weissenborn (2006) investigated German four-year-olds' intonational means of narrow focus marking. The children looked at comic strips consisting of three pictures. The pictures were accompanied by a pre-recorded story. In the last picture, a question occurred, followed by the answer, which constituted the target sentence. The children's task was to repeat the target sentence for a toy rabbit which they were told had trouble hearing. An example of the text that came with one of the stories (English translation) is given in (26).

- (26) Picture 1: Tomorrow is Peter's and Eva's mother's birthday.
Therefore, they want to surprise their mother.
Picture 2: Eva wants to bake cookies.
Picture 3: What does Peter bake?
Peter bakes a cake. [target sentence]

In order not to cue the children into any particular intonational realization of the target sentence, the target sentences were spliced together from a list of words recorded in isolation, and the fundamental frequency of the sentence was set to 150 Hz throughout, yielding an entirely flat pitch contour. It was expected that children would repeat the sentence whilst adding to it their own prosodic structure. In their analysis, Müller et al. measured the pitch values of the focused constituents, and compared these to the averaged pitch values of non-focused constituents with the same syntactic role and in the same sentence position. It turned out that the four-year-olds realized focused elements on average with a higher pitch than non-focused elements. This held both for syntactic subjects and objects, and was true both in final and initial positions. Müller et al. conclude from these results that pre-school children are able to identify the focus of the utterance (here in reply to a wh-question) and mark it

1.3 The acquisition of linguistic markings of information structure

accordingly in their prosody, a finding which is in accordance with previous studies (Hornby & Hass, 1970; MacWhinney & Bates, 1978). Note, however, that the aggregate measure “mean pitch” does not give any information about the type of pitch accent the children used.

Wonnacott and Watson (2008) investigated how four-year-olds modulated acoustic emphasis on nouns depending on their “accessibility”.²⁵ They presented the children with short video clips that showed puppets performing actions. One trial consisted of two scenes, the second of which was the target scene. Both scenes displayed the same action, but the status of the agent in the second scene was manipulated (see Table 2): In the *New* condition, the agent was entirely new, meaning that it had not occurred in the preceding scene. In the *Given-Shift* condition, the agent had been the object (theme) in the first scene. In the *Given-NonShift* condition, the agent had had the same role (agent) as in the previous picture.

Condition	First scene	Target scene
New	The bee hit the ladybug.	The <u>giraffe</u> hit the lion
Given-Shift	The elephant hugged the lion	The <u>lion</u> hugged the bee.
Given-NonShift	The ladybug kissed the giraffe.	The <u>ladybug</u> kissed the elephant.

Table 2: Experimental conditions in Wonnacott and Watson’s (2008) experiment. The target referent is underlined.

Three acoustic measures were taken: the maximum pitch at any point in the word, the overall intensity of the entire word, and the total word duration. It turned out that children produced a significantly higher maximum F0 and a higher mean intensity in the Given-Shift and the New condition compared to

²⁵ Note that Wonnacott and Watson’s use of the term differs from how it is used in this thesis, where accessible refers to an intermediate information status between given and new. Wonnacott and Watson apply the term to all referents, but assume that they differ in accessibility, rather than in information status.

1.3 The acquisition of linguistic markings of information structure

the Given-NonShift condition. There were no differences in duration between any of the three conditions. Wonnacott and Watson suggest that the results of this production experiment reflect “children’s ability to track shifts in discourse accessibility” (ibid., p. 1100). However, three aspects limit the conclusions that can be drawn from this study. First, it seems difficult to regard the individual trials as “continuing discourse”, which the set-up was meant to convey to the children (Wonnacott & Watson, 2008, p. 1096). Eliciting isolated pairs of structurally identical sentences reduced the task to marking oppositions in a very restricted linguistic environment. Second, the children were watching the scenes together with the interlocutor, which means that there was mutual knowledge between the two. Mutual knowledge as a result of a shared visual field has been found to impact children’s morphosyntactic marking of information status (as discussed in Hickmann, 2003, p. 123 f.), so it cannot be ruled out that it has an influence also on their prosodic realizations. Third, the acoustic analysis gives only very general information about the nature of the acoustic prominence, as the shape of the pitch contours was not analyzed. Like the Müller et al. (2006) study, this experiment provides acoustic evidence that children use F0 to mark referents as new, given or focal. This is a step forward as compared to the purely impressionistic analyses of earlier studies. However, neither mean F0 (the measure used by Müller et al.) nor maximum F0 (Wonnacott and Watson’s dependent variable) provide any information about the shape of the intonation contour, or type of pitch accent – which appears to play an important role in the marking of information-structural categories in adult language. The maximum pitch can occur at different positions, leading to very different accent types. A value of 315 Hz for the maximum pitch does not indicate whether the peak occurred near the onset of the stressed syllable as for example in a H*, or towards the end, as could be the case in an L*+H accent (cf. section 1.3). What is more, the two studies did not compare the children’s performance to that of adult speakers. Thus it remains unclear to what extent the children’s productions were already target-like or may have differed from that of adults, as suggested by the studies discussed at the beginning of this section.

1.3 The acquisition of linguistic markings of information structure

Unlike Wonnacott and Watson (2008) and Müller et al. (2006), Chen (2007) was explicitly concerned with the *types* of pitch accents children use in particular contexts. In an imitation experiment, she investigated how Dutch four- to five-year-old children realize topic and focus. The children first looked at a picture (e.g., a bicycle) together with the experimenter, who would comment on the object (e.g., *Look! A bicycle!*). Then, the experimenter would ask a question about that object (e.g., *Who protects the bicycle?*). In the third step, the child had to click on a robot on a computer screen, which would provide the answer (e.g., *The fox protects the bicycle*) to the child over headphones. In a procedure comparable to that of Müller et al. (2006), the robot's answer was generated by splicing pre-recorded words from a wordlist together and evening out the pitch to make the sentence intonation entirely flat. The child's task was to repeat the robot's answer to the experimenter, who was not able to hear what the robot said. The questions asked either for the subject (e.g., *Who protects the bicycle?*) or for the object (e.g., *What does the fox protect?*). The elements that were given in the question were regarded as topics, and the constituents that were the reply to the wh-elements were regarded as focus. The phonological annotation was done according to ToDI (Transcription of Dutch Intonation) guidelines (Gussenhoven, 2005).

An adult group, who performed the same task, realized both sentence-initial topics and sentence-initial foci (e.g., *The fox protects the bicycle*) primarily with H* and H*L.²⁶ In sentence-final position (e.g., *The fox protects the bicycle*), foci were typically also realized with H*(L) accents, whereas topics were most often deaccented. Like the adults, the children produced sentence-initial topics and foci mainly with H*(L) accents. In sentence-final position, children produced focused constituents also frequently with H*(L), but even more frequently with a rising accent (probably L*+H in GToBI notation). Rising contours were also found for sentence-final topics. In addition, topics were more

²⁶ Unlike GToBI, which only recognizes H*, the ToDI system distinguishes between two kinds of accents, H* and H*L. While H* indicates only a high level, H*L indicates that the pitch is falling directly after the high target.

1.3 The acquisition of linguistic markings of information structure

often deaccented, especially in sentence-final position. The high occurrence of the rising accent is unexpected, as they were hardly produced by the adult control group. However, as pointed out by Chen (p.c.), a number of children were seeking confirmation for their answers from the experimenter – probably due to the unusual task (repeating robot utterances). It is thus possible that the children were ending their utterances with a high boundary tone as a marker of uncertainty or questioning (in the sense of “Is this what you expect me to say?”), and not using the rising accent as a means to signal the topical/focal status of the referent.

Strikingly, Chen’s finding that children used a rising contour to such a large extent is at variance with the results from the studies cited above, which found that young children have difficulties producing these contours. Unfortunately, Chen does not provide any acoustic evidence for the accent labels used or phonetic detail in general. As discussed in section 1.3.1 above, a problem with purely phonological studies like this one is that it is sometimes unclear what is “behind the labels”. The problem may be exacerbated with child data, as it cannot be taken for granted that first language learners have mastered all the details of the formal markings of the language they acquire. It would thus be interesting to investigate not only whether and when children produce certain accents, but also whether the formal properties of these accents are adult-like. As will be explained in more detail in chapter 7, it is not only pitch accent realization that has hardly been investigated; the area of children’s ability to mark prosodic phrase structure is similarly unexplored.

On the whole, the studies reviewed in this section indicate that from very early on, children appear to use prosodic means to mark information structural categories such as newness or focus. It is, however, difficult to get a clear picture of their abilities, because researchers have analyzed child intonation in different ways: Some have assessed it based exclusively on perception (Hornby & Hass, 1970; MacWhinney & Bates, 1978; Wieman, 1976). Others have looked at it phonetically, using acoustic information, but did not consider the phonological aspect (Müller et al., 2006; Wonnacott & Watson, 2008). Lastly, studies interested in the phonology of child intonation have not used quantitative acoustic evidence to substantiate their claims about accent types

1.4 Research questions

(Chen, 2007; Local, 1980). It seems that a lot more could be learned about child intonation if these different levels of analysis were combined. In addition, it is unclear to what extent results that were obtained by imitation or picture description tasks generalize to children's spontaneous use of intonation in a linguistically richer environment.

1.4 *Research questions*

This brief literature review has shown that the linguistic marking of categories such as information status and pragmatic role of discourse referents has long attracted a great deal of attention (sections 1.1 and 1.2). However, with respect to first language acquisition, the research has largely focused on the morphosyntactic form of referring expressions. Intonation has hardly been studied, and the literature that exists gives only a fragmented picture. Even less is known about German. In particular, the following questions are still unanswered:

- How do German children mark topic-hood intonationally?
- How do German children mark information status intonationally?
- How do they differ from adults in these two respects, both phonologically and phonetically?

These questions are addressed in part II of this dissertation. However, from the studies discussed in this chapter it has also become clear that there are unresolved issues regarding the methodology of child and adult intonation research. Since children are mostly exposed to spontaneous adult intonation, this is expected to be their target model. At the same time, adult intonation is primarily studied by reading, implicitly assuming that reading intonation and spontaneous intonation are the same. But is this really the case? The first important methodological question that will be addressed in part III of the dissertation is therefore:

1.4 Research questions

- Is intonation in read speech comparable to intonation in spontaneous speech?

Furthermore, we have seen that there are difficulties associated with phonological labeling. Part III will therefore concentrate also on the following question:

- How can one obtain empirical support for prosody annotation (i.e., types of pitch accents and phrase boundaries)?

2 Design of the studies and outline of the dissertation

2.1 *The corpus*

The key study (presented in chapter 4) investigates the intonational marking of information status of discourse referents by German children and adults. To this end, a corpus of picture story-based narrations was created. Four additional studies draw on the same corpus of speech data elicited in that study. To give an overview, this chapter describes the general rationale behind the design of the study. More information on particular aspects of the design that are specifically relevant for the various studies, such as the exact design of the materials, will be provided in the respective chapters.

Speakers of three age groups participated in the study: five-year-olds, seven-year-olds, and adults. The two child age groups were chosen because the research reviewed in the preceding chapter suggests that some important developments take place in this age range: the appropriate morphosyntactic marking of new discourse referents does not seem in place before the age of seven, and there is evidence that certain aspects of intonation (e.g., pitch control in rising accents) are not yet adult-like in five-year-olds. We may expect that there are differences between the three age groups also in the intonational marking of information-structural categories.

In order to obtain natural data, speech was elicited by means of a picture story telling task. Unlike with free, unrestricted conversation as it is for instance found in the files in the CHILDES database (MacWhinney, 2000), picture stories make it possible to achieve some control over what is being talked about, such as the discourse referents that occur and their information status. At the same time, this design allows the participants to speak freely and to construct spontaneous discourse context in a natural way, and does not constrain their utterances as, for instance, imitation tasks or reading experiments (for adults) do. Picture stories were used rather than animated films, because they pose a smaller memory load for the child (Hickmann, 2003; Pratt & MacKenzie-Keating, 1985): With static pictures, the to-be-narrated content

2.1 The corpus

remains visible until the child has finished her narration of a given scene (Hickmann, 2003, p. 180).

Importantly, the addressee was not able to see the pictures. By avoiding shared visual space, it was ensured that the utterances are not influenced by assumed shared knowledge of speaker and hearer. For example, it may be the case that a new referent is not intonationally marked as such because the speaker assumes that the referent is not new for the hearer because it is already visually given. Colored picture sequences were created, which depicted simple stories. Typically, one main character (a human, an animal, or a cartoon character) experienced some kind of “adventure”. The plots of the stories were simple in order to be appropriate for the youngest age group. The background of the pictures (landscapes, rooms etc.) was kept simple and did not change much in the course of the story. The aim of this was to encourage participants to focus on the actions taking place (i.e., creating a story plot) and to discourage picture descriptions. An example of one story is given in Figure 7 below.

2.1 The corpus

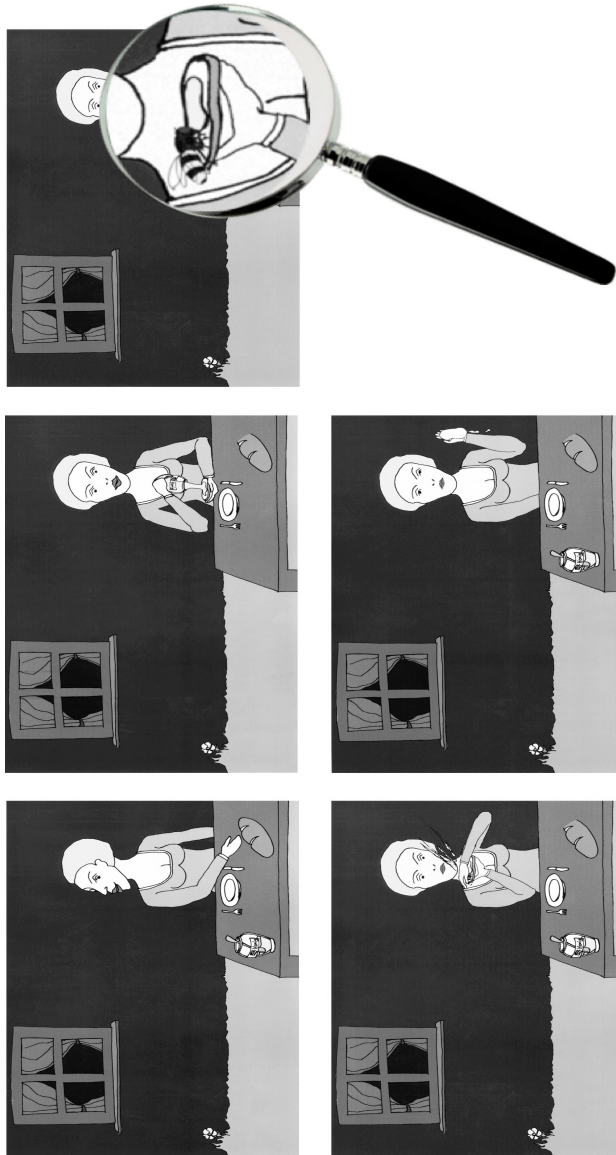


Figure 7: Example of a picture story involving a human protagonist and an animal (bee).

2.1 The corpus

The stories were presented picture-by-picture, as previous studies have found that with simultaneous presentation of all pictures children tend to focus only on the last picture (Emslie & Stevenson, 1981, p. 317). The number of pictures in a story varied from five to eight. This was done to prevent participants from anticipating the end of a story, as this could have had undesirable effects on their intonation. It has been suggested, for instance, that speakers may use features such as creaky voice to indicate the end of a larger discourse segment (Hirschberg, Litman, Pierrehumbert, & Ward, 1987; Kreiman, 1982; Lehiste, 1979). The (anticipated) end of a story would be a likely place to produce these demarcation cues. However, in particular since such non-modal phonation is difficult to deal with in intonation analysis, which is why the attempt was made to reduce the predictability of the end of each story to circumvent this problem.

Because the game-like format that proved suitable for the children would have been difficult to be convincingly used with the adult group, the procedure with which the narrations were elicited was slightly different for the children and the adults. The children were told that they were going to play a “story-telling game”. One experimenter (Experimenter A) told the children that she wanted to test the other experimenter’s (Experimenter B) memory, and that she would need the children’s help for this. The advantage of convincing the children that they are in the role of the tester rather than the “testee” is that they feel more at ease in the situation and do not have to fear that they are being somehow evaluated, a problem that is often encountered when testing school children in particular (Ginsburg, 1997, p. 19). The children were asked to tell Experimenter B several stories that Experimenter A had prepared in a “special picture book”. This picture book was a touch screen, which was used instead of a traditional paper-based book to avoid the rustling noises that usually occur when pages are turned. An additional positive side effect of using a touch screen was that it was a novelty for the children, which caught their attention and kept them engaged throughout the experiment. Children were asked to describe the events they saw and to explain everything “well”, so that Experimenter B could easily imagine what was happening in the stories and remember them. Care was taken not to use the instruction “describe

what you see” to avoid picture descriptions. Throughout the experiment, Experimenter B gave feedback to the children (*uh-huh, I see* etc.) to indicate that she was listening attentively to their narrations. The recordings, including the utterances from the experimenters, were subsequently fully transcribed orthographically using ELAN (version 3.4.0, © 2001-2008 Max Planck Institute for Psycholinguistics).

The adult participants were not engaged in the same game, but were instructed that they were going to see a number of short picture stories, which they should describe in such a way that another person listening to the recording would be able to retell the story without seeing the pictures. It was explained that the stories were simple because they had originally been constructed for children. In order to make the situation more natural, the experimenter acted as listener, giving feedback throughout the narration comparable to the feedback given to the child participants.

2.2 Outline

The data collected in the course of these experiments is subsequently analyzed from different perspectives. In the following, I will briefly describe how the studies in this dissertation are designed to answer the research questions raised in section 1.4 above.

In part II of the dissertation, I look at the relationship between certain information-structural categories (topic-hood, information status) and intonation for the three age groups. Chapter 3 investigates the intonation of topical referents in the speakers' narrations. I analyze full subject noun phrases in the so-called pre-field, that is, the position preceding the finite verb in sentences such as *Die Frau gibt dem Mann einen Apfel* (lit. 'the woman gives the man an apple'). In read speech, these types of constituents are realized with rising accents (esp. L+H* and L*+H). If this is the case in natural speech as well, adult-like marking could pose a problem for children, as they reportedly have difficulty with pitch control in rising accents. In the literature it has furthermore been suggested that contrastive topics may receive special marking. Therefore, whether contrast (defined as a change of subjecthood) has an im-

2.2 Outline

impact on the choice of pitch accent, or the phonetic realization of accents is explored.

Chapter 4 investigates the marking of information status, more specifically, the marking of activation. The discourse referents in the picture stories were manipulated to have one of three information states: new, given, or accessible. Here we look at how the global acoustic prominence variation researchers have observed in earlier child studies (Wonnacott & Watson, 2008) translates into different pitch accent types. Of particular interest is also whether we can find corroborative evidence for H+L* as a marker of accessibility in adult speech, and if so, to what extent this has been acquired by the children. In addition, this chapter examines the use of boundary tones to see how speakers of different ages handle the discourse-structuring function of intonation, and how it interacts with information status marking. As in chapter 3, whether and how children and adults differ in their phonetic realization of different pitch accent types is also investigated.

The last chapter of part II, chapter 5, examines children's ability to differentiate identifiability from activation. A subset of the picture stories elicited the production of well-known characters like *der Weihnachtsmann* ('Santa Claus'). Unlike the brand-new referents investigated in chapter 4, which were unidentifiable, these referents are new to the discourse, but identifiable for both speaker and hearer (inactive). If children have grasped the distinction between identifiability and activation, they should mark brand-new and inactive referents in the same way.

In part III of this dissertation, I concentrate on methodological issues. Chapter 6 presents the results of an experiment investigating intonation differences between spontaneous and read speech. Adult speakers read out prepared texts that were presented along with the same picture stories that were also used to elicit the corpus. The intonation in reading mode is directly compared with the intonation in spontaneous mode on the same material. This comparison can inform us about the extent to which results from reading experiments are representative for information status marking in natural speech, and whether they can be considered an appropriate target model for children.

Chapter 7 reports a perception experiment, which investigates whether judgments from untrained listeners can be used to analyze child prosody. Adult participants who have no experience with prosodic analysis listen to excerpts from the corpus. They indicate where they perceive phrase boundaries in these utterances on a written transcript, and the agreement among the labelers across the different age groups is determined. The levels of agreement are compared to those obtained in similar studies conducted on scripted and adult speech, and the usefulness of this methodology is evaluated. In addition, the results of the experiment are discussed in connection with children's ability to reliably mark prosodic structure.

In the last experimental chapter, chapter 8, we return to the issue of finding acoustic evidence for prosodic labels. In this study, nuclear contours from the corpus are modeled mathematically using third-order Legendre polynomials, improving upon the approach by Grabe, Kochanski and Coleman (2007) used for read-out English. Whether contours that were labeled differently also differ significantly from each other in their mathematical description is tested. Furthermore, how information about the relative alignment of F0 minima and maxima with the stressed syllable can be derived from the mathematical descriptions in order to make this approach more compatible with more linguistically oriented labeling schemes is explored. The last section of chapter 8 illustrates how polynomial modeling may be used to investigate age-related differences in pitch accent realization.

Chapter 9 concludes the dissertation with a summary of the main findings of chapters 3 to 8 and possible directions for future research.

Part II: Empirical Studies

Introduction

This part of the dissertation presents a set of studies investigating the relationship between certain information-structural categories and intonation in the narrations of children and adults. The first study in chapter 3 looks at the marking of topic-hood. Chapter 4 studies the marking of activation of discourse referents, and chapter 5 examines whether speakers of all age groups differentiate between the morphosyntactic marking of identifiability and the intonational marking of activation.

3 The prosodic marking of topical referents in the “Vorfeld”²⁷

3.1 Introduction

In various accounts of intonation, as well as in work on information-structure, it has been claimed that there are special accent types or intonational patterns that mark *topic* (e.g., Büring, 1997; Jackendoff, 1972; Steedman, 2000, 2007). Recently, researchers have begun investigating this issue empirically for German (Braun, 2005, 2006; Mehlhorn, 2001). These studies report that topical constituents are preferably realized with rising accents. Furthermore, German speakers modify these accents phonetically to signal *contrastiveness*.

Given the prominent place that the concept of topic and its intonational marking have been given in the literature, the acquisitional aspect has received surprisingly little attention. Using the corpus of narrations described in chapter 2, this study examines the types of pitch accents five- and seven-year-old children use to mark contrastive and non-contrastive topical referents, as well as the way these accents are realized phonetically, and compares their performance to that of adult speakers. The first part of this chapter briefly recapitulates the notion of topical referents that was presented in the background, introduces the concept of contrast, and describes how these two notions have been defined in the present investigation (section 3.2.1). Key theoretical claims and empirical findings on intonational topic marking are summarized once more (section 3.2.2). Based on this review, a number of predictions are derived. The next section provides details of the item selection and annotation. Section 4 presents the data analyses and discusses their results. It first addresses the question of whether there are any age-related differences in the types of pitch accents used for topic marking, and whether contrast affects the choice of pitch accent type. In a second step, possible influences of

²⁷ A version of this chapter has appeared as De Ruiter, L.E. (2009), The prosodic marking of topical referents in the German ‘Vorfeld’ by children and adults, *The Linguistic Review* 26(2-3) (pp. 329-354).

3.2 Background

age and contrast are explored at the phonetic level. In the final section, I turn to the interpretation of the results.

3.2 *Background*

3.2.1 **Topical referents and contrast**

The notion of topic is one of the many extensively debated concepts in the literature on information structure, and there is no generally accepted definition of the term. Consequently, there are diverging views on the defining features of topic. While purely structural definitions have been put forward (e.g., the element in sentence-initial position, Halliday, 1967b), most theorists have attempted to define topic in functional terms, often with reference to the classical definition given by Hockett: “[T]he speaker announces a topic and then says something about it” (Hockett, 1958, p. 201). This notion of “aboutness” has been adopted by most linguists (e.g., Kuno, 1972; Reinhart, 1981). However, others have argued that there are constructions that have been considered to be topic-comment structures in spite of not displaying this feature (e.g., *Körperlich geht es Peter gut*, ‘Physically, Peter is well’, from Jacobs 2001, p. 655).

Another issue concerns referentiality. It has been said that topics have to be referring, or entity-denoting (e.g., Gundel, 1999; Jacobs, 1997), that is, non-specific entities are not allowed in topic constructions (Krifka, 2007b). Others, however, see topics more as “scene-setting” expressions (e.g., Chafe, 1976), which would also include spatio-temporal adverbs (e.g., *there, yesterday*). There is also disagreement on whether topics have to be given or identifiable for the hearer, as suggested by Lambrecht (1994) or Vallduví and Engdahl (1996), for instance. However, others have suggested that new referents can serve as topics: they can be introduced to the discourse and serve as topics at the same time (e.g., Frey, 2004; Krifka, 2007a).

This cursory treatment of the topic notion illustrates that there are many controversial issues, and an in-depth discussion would be beyond the scope of this study. Since the focus of this chapter is an empirical investigation of topic marking by different age groups instead of a theoretical treatise, I have

3.2 Background

adopted a rather strict criterion in order not to include problematic cases. The analysis focuses on *aboutness topics*. The diagnostic used was a question test: I only included referents that had been previously mentioned (given) and were embedded in a sentence that could felicitously be uttered as an answer to the question “What about X?”. Further details about the item selection are provided in section 3.4.1.

A number of syntacticians have assumed that the prefield (i.e., the position preceding the finite verb) is the typical topic position in German, and have sometimes even claimed that topics *have* to be situated in the prefield in German, but there are other views as well (Frey, 2004). In contrast, not every prefield constituent is necessarily a topic, as in (27):

- (27) Vielleicht will die was malen.²⁸
 perhaps wants she sth. draw
 ‘Perhaps she wants to draw something.’

This is not a topic, because it cannot be said that the sentence is about “perhaps”.

Importantly, topics can be contrastive and non-contrastive. A topic may be called contrastive if it evokes/establishes a contrast to another (often semantically related) element, as in the theoretical example in (28). The topic constituents are underlined.

- (28) Bananen find ich lecker, aber Äpfel nicht.
 bananas find I tasty but apples not
 ‘Bananas I like, but not apples.’

A necessary condition for a topic to be contrastive is the presence of alternatives, which are often left implicit (Büring, 1997). A problematic aspect of the

²⁸ This and the following examples, including possible grammatical errors, are taken from the corpus of picture-elicited narratives, described in Chapter 2 (and in more detail in Chapter 4). Example (27) was uttered by a five-year-old.

3.2 Background

literature on contrastive topics is that they often use intonation to define or identify them (Büring, 1997; Jacobs, 1997). If one's aim is to find out when and how speakers use intonation to mark contrastive topics, this method would lead to circularity. A different approach was taken by Braun (2006; also 2005), who applied a more narrow definition of contrast: "contrast is established between two propositions of comparable referents [...]" (Braun, 2005, p. 46). In the present study, a similar but less strict working definition was used. I compared the realization of topical referents from cases in which the proposition was about the *same* referent as in the preceding sentence, with cases in which the proposition was about a referent that was *different* from that in the preceding sentence. More specifically, a topical referent was considered potentially contrastive if there was a change in subjecthood, that is, if the referent in question was not already the subject in the preceding sentence. (Only subject noun phrases were analyzed. The reason for this is discussed in section 3.4.1.) The referent may sometimes have occupied a different syntactic role in the preceding sentence (29), or may not have been mentioned in the preceding sentence at all (30).

- (29) Da fotografiert die das Kamel. Und der Kamel bückt sich.²⁹
there takes photo of the the camel. and the camel stoops itself
'There she's taking a photo of the camel. And the camel stoops
down.'
- (30) Da is das Kamel drin. Und die Frau guckt so
nach da.
there is the camel inside and the woman looks so
to there
'There's the camel in it. And the woman looks over there like
that.'

²⁹ Example (29) was produced by a five-year-old; example (30) was produced by a seven-year-old.

Change of subjecthood is clearly a broad criterion. Other studies that have looked at contrastive topic intonation have used much stricter definitions (Braun, 2005, 2006; Mehlhorn, 2001, see section 3.2.2). However, these studies were reading experiments that allow for control of the linguistic contexts in which the topics occur, whereas this investigation is analyzing natural speech in which the thematic and syntactic structures are more variable. Still, each of the picture stories always contained only two animate referents, which makes the respective other referent a plausible alternative that can evoke a contrast.

3.2.2 Topic marking in adult language³⁰

The three accent types most relevant for the present discussion are H*, L+H*, and L*+H, which were introduced in chapter 1. For the reader's convenience, they are briefly described again here. The first accent, H*, is described as a peak accent, which may be preceded by a shallow rise. The accented syllable is perceived as high. The second accent, L+H*, is characterized by a steep rise in the accented syllable, whose target may only be reached in the following syllable, but the accented syllable is still perceived as high. The third accent, L*+H, is also characterized by a steep rise, but it differs from L+H* in featuring a local pitch minimum in the accented syllable and sounding clearly low. The rise starts only later in the accented syllable. The three accent types are illustrated in Figure 8.

³⁰ For a detailed discussion on previous literature on the prosodic marking of topics in German, the reader is referred to Braun (2005).

3.2 Background

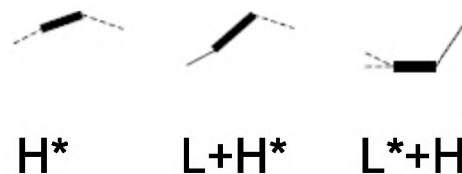


Figure 8: Schematic representation of the accent types H*, L+H* and L*+H in the GToBI inventory, taken from Grice and Baumann 2002, p. 12. Thick black lines indicate the accented syllable, solid black lines indicate compulsory pitch movement, and dotted lines indicate optional pitch movement.

For English, Steedman (2000, 2007) suggests that sentence-initial topics are marked with L+H* or L*+H, and has extended this claim to topic marking in German as well (Braun 2005, p. 55, footnote 5). Steedman assumes that L+H* accents are used for “agreed” topics, whereas L*+H accents are used “non-agreed” topics. The distinction agreed/non-agreed seems comparable to that of contrastive/non-contrastive. Büring (2003, 2007) claims that contrastive topics in English are realized with H* or L+H*, but does not specify in which way contrastive topics differ intonationally from non-contrastive ones.

For German, most claims are about contrastive topics. The accent pattern for these so-called “I-topics” (Jacobs, 1982) has been described as a fall-rise movement, referred to as a “root contour” (Jacobs, 1997), and accordingly represented with the root symbol “√”. In AM terms, the root contour would be most adequately represented as a L*+H accent (see Jacobs 2001, p. 92). Mehlhorn (2001) proposes that (in the absence of contrast) topicality can be expressed with a “simple lexical accent without prominence” (Mehlhorn, 2001, p. 39, my translation). It is not clear what this term refers to. It may either mean simply lexical stress, and thus imply that non-contrastive topics may not receive an accent at all; or it means that there is some kind of sentence-level stress which is phonetically distinct both from lexical stress and pitch accent, a notion which I find difficult to follow. Contrastive topics, on the other hand, Mehlhorn assumes to be realized with the root contour.

3.2 Background

In addition to the type of pitch accent used, there has also always been an interest in potentially functional differences in accent *realization*. Dimensions of interest have been the horizontal alignment and the vertical scaling of pitch movements (i.e., the height of the pitch peaks and the depths of the troughs). Ladd (1983) proposes a phonological feature [raised peak] to describe the increased pitch height in contrastive accents. In later accounts (Ladd, 1996), this “emphatic peak raising” has been analyzed as being gradient in nature, rather than categorical.

Regarding the empirical investigation of these theoretical claims, so far only two studies have addressed topic marking in German. Mehlhorn (2001) conducted a production experiment in which subjects were asked to read out question-answer pairs. The questions provided the information structural context for the answers. These target sentences were either contrastive (“I-topics”) or non-contrastive (“neutral”). The target words (topics) occurred in both conditions. The acoustic characteristics of the target words, which were always in sentence-initial position, was analyzed and compared. Unfortunately, Mehlhorn does not give details of her analysis and does not discuss the types of accents that were used. However, she does report that there was an F0 rise on the lexically stressed syllable of the topic constituent in all test sentences. This suggests that all topics were accented, counter to Mehlhorn’s own assumption that the perceived prominence in neutral topics is due to lexical stress only. With respect to the phonetic realization of the rising accents, contrastive topics were realized with a greater excursion and steeper F0 rise on the accented syllable. In addition, the syllable duration was longer in the contrastive contexts.

The most extensive work on the prosodic marking of topics in German has been carried out by Braun (2005, 2006). Braun studied both the production and the perception of contrastive and non-contrastive topics (she uses the term *themes*, following Steedman, 2000). In her production experiment, participants read target sentences embedded in short paragraphs, which provided contrastive or non-contrastive contexts for the target sentences. The same sentence was produced in both context types. This design allowed for a direct comparison of the two conditions for each subject individually. Braun

3.3 Topics in child language

found that all topics were realized with the prenuclear accents L+H* and L*+H. Furthermore, the two accent types were equally distributed across both conditions. Contrastive and non-contrastive topics were thus not differentiated phonologically. On the phonetic level, however, subjects distinguished contrastive topics from non-contrastive ones by using higher peaks, a greater excursion and later alignment of the peaks. Both Mehlhorn's (2001) and Braun's (2005, 2006) findings support previous claims in the literature that topics in German are realized with rising accents. However, they do not support the hypothesis that contrast is signaled by a particular pitch accent type, as suggested by Steedman (2000, 2007). Rather, they found that contrastiveness is expressed by gradient phonetic modification of an accent.

3.3 *Topics in child language*

As discussed in the background, studies on topic encoding in the syntactic domain report that children under three years of age show appropriate usage of word order and syntactic constructions to mark topic-hood (e.g., Scollon, 1979; De Cat, 2003). In the prosodic domain, researchers have mainly looked at *focus* (Hornby & Hass, 1970; MacWhinney & Bates, 1978), with results showing that English- and German-speaking children (at least from four years onwards) accent focused constituents. However, few studies have addressed the prosodic marking of topics in child language. One study that was explicitly interested in topic marking was Chen's (2007) experiment mentioned in chapter 1, which used an imitation paradigm to investigate Dutch children's intonation to mark topic and focus. Chen defined topics as the given elements in the answer to a narrow focus question (e.g., *What does the fox protect? – The fox protects the forest*). She found that for sentence-initial topics, children resembled adults in that they used mostly H* accents, but differed from adults in that they produced deaccented constituents more often. However, Chen's findings are not directly applicable to the present study, as her topicality was defined by the question context, which cannot be applied here. In Chen's study, topicality was equated with givenness: The given referent was automatically defined as the topic of the sentence. Furthermore, as pointed out in chapter 1, the children in Chen's study were given the sentences they were to produce in the form of a synthetic stimulus that had no intonation,

3.3 Topics in child language

the form of a synthetic stimulus that had no intonation, which makes these utterances very different from natural speech.

Wonnacott and Watson's (2008) experiment was not explicitly concerned with topic marking, but it bears some relevance to the present investigation, because it also analyzed the prosody of previously mentioned referents in sentence-initial position, which is why it will be briefly repeated here. Children described two consecutive scenes from short video clips that showed puppets performing certain actions. Both scenes displayed the same action, but the status of the agent in the second scene was manipulated: In the Given-Nonshift condition, the agent had had the same role in the previous picture. In the Given-Shift condition, the agent had a different role in the first scene (e.g., experiencer) than in the second. The Given-Nonshift and the Given-Shift conditions are to some extent comparable to the contrastive and non-contrastive topics as they have been operationalized here. Children produced a significantly higher maximum F0 and a higher mean intensity in the Given-Shift condition compared to the Given-Nonshift condition, indicating that they use prosody to mark changes in agenthood. But, as pointed out in chapter 1, we do not know which types of pitch accents the children used, and there is also no comparison of the children's productions with that of adults.

It is also worth repeating that acoustic investigations of children's intonation production have found that five-year-old children may have difficulty imitating nuclear rising contours (Loeb & Allen, 1993; Snow, 1998), and that they produce nuclear accents with a narrower pitch range (excursion) and with a slower speed of pitch change (slope) than adults (Snow 1998). These findings are relevant for the present study, given that prototypical topical accents in German are rises. Note, however, that previous findings concern nuclear accents, whereas the present investigation is concerned with prenuclear accents.

To sum up, past studies have found that four-year-old children already use intonation to signal information-structural aspects like focus, and that they mark changes in subjecthood prosodically. There is also evidence that their phonetic realization of pitch accents may not yet be adult-like, and that

3.4 Method

in particular the realization of rising accents may pose some difficulty. On the basis of the reviewed literature, I put forth the following hypotheses:

1. In adult speech, topical referents in the prefield are realized predominantly with L+H* and L*+H.
2. The proportion of accents with a steep rise (L+H*, L*+H) increases with age.
3. The F0-excursion in the accents increases with age.
4. The speed of pitch change (slope) increases with age.³¹
5. In sentences with a change of subjecthood, the accents are realized with greater excursion, steeper slope and later aligned peaks than in sentences where the agent (subject) is maintained from the preceding sentence.
6. Change of subjecthood leads to a preference for L*+H accents compared to L+H* accents in contexts in which subjecthood remained constant.

3.4 Method

The analyses are based on the corpus of picture story-elicited narrations described in chapter 2. As mentioned there, the materials were designed for the study of information status marking, which is presented in the next chapter (chapter 4). To avoid repetition, the following paragraphs will highlight only the main aspects relevant to the present study.

Twenty-nine five-year-olds (range = 4;11–5;11, mean = 5;05, 11 boys, 18 girls) and twenty-six seven-year-old children (range = 6;11–7;08, mean = 7;03, 9 boys, 17 girls) participated in the study. The children were recruited from kindergartens and primary schools in the *Niederrhein*-area of Germany (north-west of Germany). Informed written consent was obtained from caregivers or parents. Twenty-eight adult speakers (mean age 23 years, 5 male, 23

³¹ Assuming that speaking rate remains constant or (more likely) even increases with age (e.g., Boutsen & Hood, 1996; Sturm & Seery, 2007), 4 follows from 3. However, for the sake of completeness, both points are listed here separately.

female) acted as controls. Most of them were undergraduate students of Psychology or Linguistics at the University of Potsdam. All participants were native speakers of German and had no reported history of speech/language or other developmental deficits. The children received stickers for their participation. The adult participants were given course credit or a financial reimbursement.³²

The children were engaged in the previously described story-telling game. They sat at a comfortable distance in front of an Elotouch ET1525L touch screen, which served as the picture book. Neither Experimenter A nor Experimenter B could see the screen. Children went through the stories at their own pace. The screen presentation was controlled with the Nijmegen Experiment Set-up (NESU) software. The sessions were sound-recorded using a Roland Edirol R-1 24bit digital Wave/MP3 recorder and a Sony ECM-950 DT microphone at 44.1 kHz (16bits precision, stereo sound). The microphone was placed at about 20 cm distance from the child. Throughout the experiment, Experimenter B gave feedback and asked questions to maintain their interests (“Wow, that is really interesting”, “And what happens then?”, etc.).

The adults were recorded in a sound-attenuated booth, but were connected with the experimenter via headphones to enable interaction. They followed the screen presentation on a computer screen outside the booth through a window. The stories were presented using Microsoft PowerPoint. The speakers were recorded using a Tascam DA-45HR 24bit digital DAT-Recorder and an Audio Technic AT 4033a/B microphone at 44.1 kHz (16bits precision, stereo sound), placed at about 30 cm distance.

³² Data from four additional children (three five-year-olds and one seven-year-old) could not be used, because the children were too shy (one child), too distracted (two children), or did not understand the stories (one child). Another child later turned out to be bilingual, so her data was discarded as well. Two recordings from adult speakers had to be excluded because of equipment failure.

3.5 Item selection and annotation

3.5 *Item selection and annotation*

3.5.1 Item selection

Sentences were selected in which the prefield position was occupied by a noun phrase denoting a referent that had been introduced before, and if they were a felicitous answer to the question *What about X?*. This excluded sentences with sentence modifiers or adverbials (e.g., *unfortunately, then*), and utterances like (31) where a referent in preverbal position is realized with an indefinite article:

- (31) Ein Biber kommt vorbei.
a beaver comes along
'A beaver comes along.'

Four additional selection criteria were applied for practical reasons. First, only full noun phrases were looked at. Of course, topic referents can take on various lexical forms, such as full noun phrases (NPs), pronouns and null forms. However, the intonational phenomena of interest here can best be observed on full NPs. Second, only animate referents were analyzed. Since animacy has been found to have an impact on case marking in German child language (Drenhaus & Féry, 2008), it may influence intonation as well. Third, since children did not produce object NPs in the prefield, I restricted the analysis to subject NPs to keep the data comparable across age groups. Finally, items with bad recording quality, creaky voice or laughter were excluded. Altogether, 271 utterances were selected, of which 169 came from the adult group, 74 from the seven-year-olds, and 28 from the five-year-olds. The low number of items in the youngest age group is primarily due to the fact that the children often produced connectives in combination with temporal adverbs in the preverbal

3.5 Item selection and annotation

position (e.g., *und dann*, ‘and then’). Table 3 gives an overview of the frequency of these constructions³³ for each age group.

	Age group		
	5 years	7 years	adults
	N	N	N
und dann ‘and then’	364	336	66
und jetzt ‘and now’	284	315	20
animate subject NPs	28	74	169

Table 3: Frequency of connective+temporal adverb constructions and topical referents (animate subject NPs) in sentence-initial position by age group.

3.5.2 Annotation

The utterances were annotated and analyzed using Praat version 5.0.35 (Boersma and Weenink 1992-2008). For the phonological analysis, I labeled the topical NPs intonationally following GToBI guidelines (Benzmüller & Grice, 1997). Furthermore, phonotactic and segmental information were extracted from the corpus and included in the statistical analysis (see section 3.4.3).

For the phonetic analysis, the data was annotated both on the segmental and on the suprasegmental level. The intonational phrase containing the topical NP was segmented at the level of the syllable using information from a

³³ Recall that on several occasions the experimenter (listener) asked broad focus question like “And what happens then?”. These questions often prompted a response beginning with “and then” on part of the speakers. In the count reported here, all these cases have been already subtracted from the overall number.

3.5 Item selection and annotation

wide-band spectrogram, and the onset and offset of the lexically stressed syllable were marked. The pitch contour was corrected manually for artifacts produced by the pitch tracking algorithm (doubling and halving errors). Then position and value of local F0 maxima and minima were determined manually for the three predominant accents H*, L+H* and L*+H. The domain in which these landmarks were set consisted of the stressed syllable, its preceding and following syllable.³⁴ An example is provided in Figure 9.

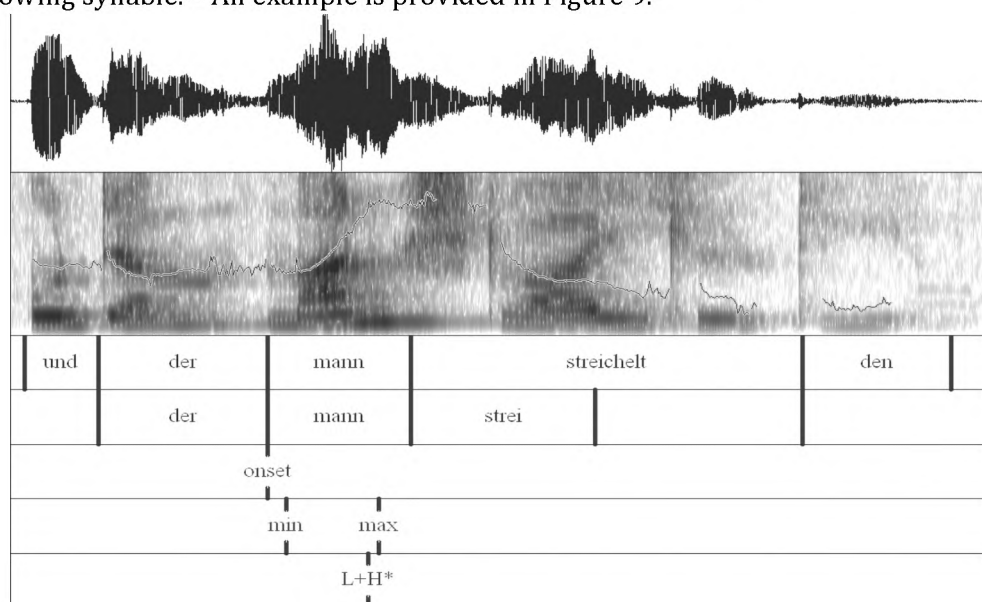


Figure 9: The sentence *Und der Mann streichelt den*, ‘And the man is stroking him’, uttered by a five-year-old. The annotation contains the onset of the stressed syllable, pitch minimum (min) and maximum (max), and the pitch accent type according to GToBI.

³⁴ Only seven cases of accents other than H*, L+H* and L*+H were found, therefore these were not annotated for phonetic analysis. Where an item was labelled as deaccented, the maximum and the minimum within the three-syllable-domain were measured.

3.5 Item selection and annotation

From these landmarks, a number of measures were derived:

- the excursion of the rise (F0 max-min; in semitones³⁵ (st))
- the duration of the rise³⁶ in seconds (sec)
- the slope of the rise (st/sec)
- the position of the F0-maximum (% into the stressed syllable)
- the position of the F0-minimum (% into the stressed syllable)

Accent realization in German has also been found to be influenced by syllable structure (Grabe, 1998a, b).³⁷ In order to be able to detect such effects, each item was additionally coded with respect to its syllable structure:

- the number of syllables of the word
- onset voiced/unvoiced
- coda voiced/unvoiced

I further annotated the occurrence of a pause following the topical NP, and its length (in milliseconds). Finally, a change in subjecthood was flagged as a potential source of contrast.

3.5.3 Consistency check of accent labels

Before carrying out the phonological analysis, I first wanted to make sure that the accents to which different labels were assigned do indeed represent the different GToBI categories. As discussed in section 1.3.1 above, consistent prosodic annotation is often difficult. In their agreement study, Grice et al. (1996) had found that in particular the pairs L+H* and H*, and L+H* and L*+H were

³⁵ The semitone scale is a logarithmic transformation of the (linear) Hertz scale. It is a scale which provides a better representation of the perception of differences in pitch than the Hertz scale (Nolan, 2003).

³⁶ The duration was measured as the time between the F0-minimum and the F0-maximum.

³⁷ Grabe (1998a, b) found that nuclear falls preceding a low boundary tend to be truncated when the amount of sonorant material is reduced (as in the case of unvoiced codas, such as in [jɪf], rises before high boundaries tend to be compressed.

3.5 Item selection and annotation

prone to confusion. These are the predominant accents in the present data set. Given that further analyses depend on the accuracy of the labels, it is desirable to obtain additional support for the correct application of the GToBI system.

In the labeling guidelines for GToBI, a number of distinguishing features are described to help identify the accent types (see Table 1 in chapter 1). These descriptions can be turned into testable predictions regarding their differences along a number of acoustic parameters. The parameters considered to be relevant here are: F0 excursion, slope, position of the F0-maximum and position of the F0-minimum.

I tested these predictions using the statistical method of multilevel regression modeling (Baayen 2008; Baayen et al. 2008; Bates and Sarkar 2007). Multilevel modeling allows for the incorporation of fixed factors with an explanatory value (such as accent type) together with crossed random factors such as subject or item. These linear mixed-effect models (LME) have a number of advantages over traditional methods like ANOVA, such as robustness with respect to missing data (Baayen 2008, p. 290). All analyses were performed using the statistical software R (R Development Core Team, 2008).

I tested whether pitch accent type (henceforth ACCENT) is a significant predictor for the dependent variables excursion (EXC), slope (SLOPE) and alignment of the F0 maximum (MAX) and the F0 minimum (MIN). The models therefore contained the acoustic parameter in question as the dependent variable and ACCENT as fixed factor. In order to detect possible effects of the segmental make-up of the words, these other fixed factors were included in each model: the number of syllables of the word (SYLLNO), onset voiced/unvoiced (ONSET), coda voiced/unvoiced (CODA). If a factor or an interaction is not explicitly mentioned in the description of the individual models, it means that this factor or interaction did not have a significant effect and was therefore excluded.

In addition to the fixed factors, the following variables were initially entered into each model as random factors, but excluded if the estimated variance turned out to be effectively zero: subject (SUBJECT), word (WORD), duration of pause (PAUSE). If the explained variance was not zero, likelihood ratio tests were used to test whether the random effect parameters are justified in

3.5 Item selection and annotation

the model. All reported p-values were obtained by estimating the posterior probability of a Markov Chain Monte Carlo (MCMC) simulation with 10,000 runs.

Before looking at the phonetic differences between the three major accent types, I checked whether the perceptual impression that a number of items did not bear any accent could be backed by the phonetic measurements. When a constituent is deaccented, there should be only little pitch movement in the three-syllable domain, meaning that EXC and SLOPE should be small. In order to test this, EXC and SLOPE in deaccented items were compared to those in items labeled H*, L+H* and L*+H. The LME models confirmed that EXC in deaccented referents was significantly smaller than in referents labeled H* (2.07 st vs. 3.38 st, $p < .01$), as well as in referents labeled L+H* (2.07 st vs. 6.54 st, $p < .0001$) and referents labeled L*+H (2.07 st vs. 6.30 st, $p < .0001$). Similarly, SLOPE was significantly shallower in deaccented referents compared to referents labeled H* (-2.32 st/sec³⁸ vs. 12.49, $p < .0001$), referents labeled L+H* (-2.32 st/sec vs. 23.31 st/sec, $p < .0001$) and referents labeled L*+H (-2.32 st/sec vs. 22.88 st/sec, $p < .0001$). Having established that deaccented items differ significantly from accented items, I now turn to the systematic comparison of the three accent categories H*, L+H*, and L*+H. Excluding 7 items with other accents and 42 deaccented items, the data set subjected to the analyses consisted of 222 observations.

H* vs. L+H*

In GToBI, the accent L+H* differs from H* in that it is preceded by “a low pitch target which leads to a sharp rise in [...] the accented syllable” (Grice et al. 2005, p. 65). L+H* accents would thus be expected to have a greater pitch excursion and a steeper slope than H* accents. The LME model with EXC as the dependent variable revealed a significant effect of ACCENT: The excursion in L+H* accents was on average about 3.1 st higher than in H* accents (6.61 st vs. 3.48 st, $p < .0001$). A significant effect of ACCENT was also found for SLOPE. The slope in L+H* accents was about 10.03 st/sec steeper than the slope in H*

³⁸ The negative slope indicates that the pitch in deaccented items was slightly falling.

3.5 Item selection and annotation

accents (14.20 st/sec vs. 24.23 st/sec, $p < .0001$). The accents labeled as H* and L+H*, respectively, differ thus significantly from each other in both excursion and slope in the expected direction.

L+H* vs. L*+H

According to GToBI guidelines, the accents L+H* and L*+H are both characterized by a steep rise, but they are said to differ in the position of the pitch minimum and pitch maximum. In L+H*, the minimum is located before or at the beginning of the stressed syllable, and the rise occurs typically within the stressed syllable, peaking late in the accented syllable. In L*+H, the minimum is supposed to be clearly within the accented syllable, with the rise beginning only very late in the syllable. The maximum is typically reached only on the first post-tonic syllable, or even later. The LME models confirmed that in L*+H accents, MIN occurred on average 28% later than in L+H* accents (34.4% into the accented syllable vs. 6.3% into the accented syllable, $p < .0001$). While MIN in L+H* occurred at the beginning of the accented syllable, it is clearly located within the accented syllable in L*+H accents. A significant difference between the two accent types was also found for the position of the peak. In L*+H accents, MAX occurred about 34.8% later than in L+H* accents (118.2%³⁹ into the accented syllable vs. 83.4%, $p < .0001$). As predicted, the peak in L+H* occurred late within the accented syllable, while it was pushed beyond the end of the accented syllable in L*+H accents. Thus in accents labeled L*+H, both the pitch minimum and the pitch maximum were aligned significantly later than in accents labeled L+H*. There was also an effect of SYLLNO: In disyllabic words, MAX was realized significantly later than in monosyllabic words (e.g., for L+H* accents: 102.9% vs. 83.4% into the accented syllable, $p < .05$). This can be explained by the fact that if more sonorant material is available (as is the case in disyllabic words), speakers can realize the pitch peak later.

The preceding analyses demonstrate that the accents that were labeled as H*, L+H* and L*+H differ from each other along a number of acoustic param-

³⁹ A value larger than 100% indicates that the F0 maximum occurred after the offset of the accented syllable, that is, within the post-accentual syllable.

ters as predicted by the GToBI assumptions. It is therefore assumed that the labels have been applied consistently⁴⁰, and all annotations were retained.

3.6 *Results and discussion*

3.6.1 Phonological analysis – choice of pitch accent type

This section addresses three questions. First, which pitch accent types⁴¹ did German children and adults in this study use to mark topical referents in the prefield? Second, are there any differences across age groups? Third, does CONTRAST influence the choice of pitch accent type? The data set used for the analyses comprised 264 items (all items minus the ‘other’ category). The overall distributions of pitch accent types for each of the three age groups are displayed in Table 4.

⁴⁰ Note that this procedure only checks whether the labels have been assigned consistently according to the accent categories as postulated in the GToBI annotation system. As mentioned in the background section, the question to what extent these categories represent “true” phonological categories of German is a different issue altogether, a discussion of which is beyond the scope of this thesis. The interested reader is referred to Kügler (2007) and Grice and Baumann (2002) and references therein.

⁴¹ Arguably, “deaccented” is not a type of pitch accent, but for reasons of brevity I will use the term “(pitch) accent type” to also include the category “deaccented”.

3.6 Results and discussion

		Age group					
		5 years		7 years		adults	
		N	%	N	%	N	%
Accent. type	deacc.	2	7.1	11	15.7	29	17.5
	H*	9	32.1	30	42.9	27	16.3
	L+H*	15	53.6	18	25.7	57	34.3
	L*+H	2	7.1	11	15.7	53	31.9

Table 4: Distribution of accent types (excluding “other”) by age group.

There appear to be four major differences among the age groups. First, there is the low proportion of L*+H accents in both the five-year- and the seven-year group compared to the adult group. Second, there seems to be an increase in the amount of deaccented items by age. Third, the proportion of H* accents is much higher in the five- and seven-year-olds compared to the adults. Lastly, the youngest age group has used more L+H* accents than the two other groups.

These differences were analyzed statistically by performing logit mixed-effect analyses (see Bates & Sarkar, 2007; Jaeger, 2008 for the analysis used) using accent type (ACCENT) as the dependent variable, AGE as fixed factor (predictor), and SUBJECT, WORD and PAUSE as random factors. Because a multinomial analysis is not yet implemented in R, the variable ACCENT was recoded into binary variables (e.g., “H* yes/no”) and binomial analyses were performed. In addition to an overall difference in accent type distribution, the age groups may also differ with respect to the use of accent type to mark contrast. I therefore also included CONTRAST as a second predictor variable in the model. Finally, ONSET, CODA and SYLLNO were included as additional predictors to test whether the frequency of occurrence of a given pitch accent type was influenced by any of these factors.

3.6 Results and discussion

The analysis showed that most of the observed differences were significant. Although there was no significant difference between the age groups regarding the amount of deaccented referents, the assumption that the younger speakers used fewer L*+H accents than the adults was partially borne out. The difference between five-year-olds and adults approached conventional levels of significance ($z = 1.83$, $p = .06$). There was, however, no significant difference between the five- and the seven-year-olds ($p > .1$), and seven-year-olds did also not differ from adults in the frequency of L*+H ($p > .3$).

Regarding the proportion of H*, the difference between five-year-olds and adults also approached significance ($z = -1.77$, $p = .07$), and was significant between seven-year-olds and adults ($z = 3.55$, $p < .001$). The two younger age groups did not differ from each other.

Finally, the seven-year-olds did indeed produce fewer L+H* accents than both the five-year-olds ($z = 3.07$, $p < .01$) and the adults ($z = 2.40$, $p < .05$), but five-year-olds and adults did not differ from each other ($p = .13$). Unexpectedly, there was no effect of CONTRAST: None of the age groups showed a preference for a particular accent type in cases where there was a change of subjecthood. ONSET, CODA and SYLLNO did not affect the choice of pitch accent type.

The results of the phonological analysis bear out some of the predictions made in section 2.5. The adult speakers in this study realized topical referents predominantly (in two thirds of the cases) with L+H* and L*+H accents, confirming prediction 1. There was also an increase in the use of L*+H accents with age, partly confirming prediction 2. However, this age trend did not hold for L+H* accents. Finally, change of subjecthood did not lead to a preference of L*+H accents over L+H* accents, counter prediction 6. The pattern emerging from the analyses of the pitch accent type distributions is summarized in Table 5.

3.6 Results and discussion

		Pattern
Accent. type	deacc.	5 = 7 = adults
	H*	5 = 7 > adults
	L+H*	7 < 5 = adults
	L*+H	5 < 7 = adults

Table 5: Pattern of pitch accent type distribution across age groups. An “=” sign indicates that no significant difference was found between the two age groups; a “>” sign indicates that the age group left to the sign produced this accent type significantly more often than the group right to the sign; a “<” sign means that the group on the left produced the accent significantly less often than the group on the right.

The finding that the five-year-olds seem somewhat more adult-like in their production than the seven-year-olds (viz. the more frequent use of L+H*) is puzzling. At this point one might object that the existence of the L+H* category as a category distinct from H* has been disputed (Féry 1993; Grabe 1998b). The acoustic analyses, however, showed that in the present data there were clearly two distinct accents, H* and L+H*, suggesting that there is a real difference between the two age groups. In order to get some additional information about the five-year-olds’ intonation patterns in the prefield, I looked at the realizations of *und dann* (‘and then’). The five-year-olds produced 364 utterances of that type in their narrations (see section 3.3.1). From the 364 utterances, a random sample of 40 utterances was selected. As with the other data, the prefield constituents were labeled for their accentuation type according to GToBI. It turned out that the overall distribution of accentuation types for *und dann* utterances differed considerably from that of the other data. In half of the utterances (20), these prefield constituents were not accented at all. The prevailing pattern in these cases was a high boundary tone at the beginning of the phrase, followed by a steady declination toward the first accent later in the phrase. H* occurred seven times, L+H* six, and L*+H five times.

3.6 Results and discussion

The fact that the accentuation patterns from the two data sets are clearly different indicates that the *und dann* utterances seem to play a different role in discourse structuring than the topical referents – an issue which will be taken up again in the general discussion. Due to their different status, these items cannot be used to supplement the analysis of the topical referents. Note that this difference across data sets also shows that the intonation patterns in the original data set cannot solely be determined by their sentence-initial position but must also be influenced by the pragmatic role of the constituents.

3.6.2 Phonetic analysis

On the phonetic level, the same questions arise as on the phonological level: Are there any differences between the age groups? And does the presence of contrast affect pitch accent realization in any way phonetically, as has been found by Mehlhorn (2001) and Braun (2006)?

These two questions can be addressed simultaneously by using linear mixed-effect models again, directly predicting the continuous acoustic variables. If there are differences between the age groups, then the factor AGE should be a significant predictor for the four acoustic parameters EXC, SLOPE, MIN and MAX. Likewise, we should be able to observe a significant main effect of CONTRAST, if a change in subjecthood influences the phonetic realization. We already know from the consistency check of accent labels (see section 4.4.1) that ACCENT was a significant predictor for the four acoustic parameters. It therefore needs to be included in the model. The data set subjected to the analysis contained all items labeled H*, L+H* and L*+H (222 observations).

For each acoustic parameter, an LME model was initially specified with the acoustic parameter as dependent variable and ACCENT, AGE, CONTRAST, as well as SYLLNO, ONSET and CODA as fixed factors. In the initial model, all main effects and interactions were tested. Only those predictors and their interactions that were significant at $p < .05$ were maintained. All main effects that appeared in significant interactions were also maintained. The fitting procedure for the random effects is as described in section 3.4.3 above. As ex-

3.6 Results and discussion

pected, ACCENT was always a significant predictor, but did not interact with the other predictor. It will not be reported again in the individual analyses.

Excursion

The intermediate age group produced accents with a significantly smaller excursion than the adult group (e.g., in H* accents: 3.6 st/sec vs. 4.2 st/sec, $p < .01$). The difference between the seven-year-olds and the five-year-olds was marginally significant (in H* accents: 3.6 st/sec vs. 3.8 st/sec, $p = .05$). There was no difference between the five-year-olds and the adults. CONTRAST was not a significant predictor for excursion, neither were SYLLNO, ONSET or CODA.

Slope

Adult speakers produced accents with a significantly steeper SLOPE than the two younger age groups (e.g., in H* accents: 10.4 st/sec vs. 8.4 st/sec in the five-year-old group, $p < .0001$; 10.4 st/sec vs. 7.4 st/sec in the seven-year-old group, $p < .001$). The five-year-olds and the seven-year-olds, however, did not differ significantly from each other. CONTRAST was not a significant predictor for slope. There were no significant effects for SYLLNO, ONSET or CODA.

Position of the F0 minimum

Both the five- and the seven-year-olds turned out to align MIN earlier than the adults (e.g., in L+H* accents: five-year-olds at a position of 17.6% preceding the accented syllable vs. adults 6.1% into the accented syllable, $p < .05$; seven-year-olds 10.8% before accented syllable vs. 6.1% into the accented syllable, $p < .01$). The two younger age groups did not differ from each other. The analysis showed also an interaction between AGE and ACCENT. Compared to adults, seven-year-olds realized the minimum in L*+H accents later (39.0% vs. 65.1% into the accented syllable, $p < .001$). There was no effect of CONTRAST, SYLLNO, ONSET and CODA.

Position of the F0 maximum

The three age groups did not differ from each other in terms of the alignment of the pitch maximum (e.g., in H* accents mean MAX adults 62.3% into the accented syllable, mean MAX seven-year-olds 63.1%, mean MAX five-year-olds 59.83%). There was also no effect of CONTRAST. A significant effect was found for SYLLNO, paralleling the results from the consistency check of the pitch accent labels (section 3.3.3). MAX occurred later in disyllabic items than in monosyllabic items (e.g., for H*: 83.3% into the accented syllable vs. 62.3%, $p < .05$).

The results of the phonetic analyses partly confirm the developmental predictions made in section 2.5. Adults realized accents with a greater excursion than seven-year-olds, in keeping with prediction 3. However, five-year-olds did not realize accents with a smaller excursion than the older age groups, counter prediction 3. The younger age groups both produced accents with a smaller speed of pitch change (slope) than adults, confirming prediction 4. As age differences in tonal alignment have not been reported before, an unpredicted result was that the children consistently aligned pitch minima earlier than the adult group. Prediction 5 (higher excursion, steeper slope and/or later alignment of peaks in contrastive contexts) was not borne out by the data.

3.7 General Discussion

This study has investigated the prosodic marking of topical referents in the German prefield by children and adults. The first result is that the data for the adult speakers replicate the main findings from Mehlhorn (2001) and Braun (2006). In the adult group, topics were mainly realized with rising accents. As in Braun's experiment, L+H* and L*+H accents occurred equally often. The results differ from the aforementioned studies in that I also found a considerable amount of monotonal H* and deaccentuation. These patterns are not typically assumed to be topic accents in German and have so far not been attested in experimental studies. However, Hedberg and Sosa (2007) found a high proportion of deaccentuation and H* accents as well in their analysis of

3.7 General Discussion

topic accents in spontaneous American dialogues. Deaccentuation and H* accents may thus be more common as realizations of topical constituents in natural speech than in read speech.

Before we turn to the interpretation of the child data, let us first consider another striking finding, namely relative scarcity of data points for the younger age groups overall. Referents in sentence-initial topic position were generally hard to find in the child data. Even the seven-year-olds produced less than half as many (74) topical referents in the prefield as the adults did (169). Instead, the children began most of their utterances with the connective *und*, 'and' together with temporal adverbs (*dann* 'then', *jetzt* 'now'), constructions adults made much less use of (see Table 3). As remarked before, all utterances that were responses to broad focus questions like *And what happens then?*, in which such an answer would be quite natural are disregarded in these numbers. Still, the recurrent question may have cued children to assume that every utterance should be an answer to this (implicit) question, and may have prompted them to start most of their sentences with 'and then'-like constructions. However, it is unlikely that the high rate of occurrence of these constructions is a consequence of the experimenter's prompting, as the same observation has been made before in other studies that did not systematically use prompting questions like the ones used here. Hickmann et al. (1996) found that German five- and seven-year-olds produced nearly 70% of all previously mentioned referents in postverbal position due to these kinds of constructions (compared to less than 30% in the adult group). It is assumed that these constructions are initially empty discourse fillers whose function is to signal that more is to come (Berman, 1996). Later they are used to express temporal sequentiality. Von Stutterheim and Carroll (2005) assume that speakers may choose different criteria to achieve coherence in narrations – they may focus on the temporal shift; in German, adverbs indicating this shift are likely to occur in the preverbal position. Or they may organize the discourse around a topical referent, in which case this referent is located in the pre-field. The present data suggest that children prefer the first strategy, whereas adults more often opt for the second. This may be due to the fact that adults have already learned that linear temporal progression is the default in

3.7 General Discussion

narratives (see e.g., Klein's (1986) *principle of natural order* [PNO]), and that it does not always have to be specified overtly. Children, on the other hand, still focus on the chronological order of events, linking propositions together on a local level. Topical referents may therefore not be as common as in adult speech.

The data from this study suggests that *when* German children produce topical referents in the prefield, their intonation patterns are similar, but not identical to those of adults. The five- and seven-year-olds used the same set of pitch accents as the adult group, but differed in the frequency with which they used the different accent types. The children used more H* accents, which are produced with a less steep pitch rise than L+H* and L*+H, and have been considered to be the default accents in German (Wunderlich, 1988). The youngest age group hardly produced L*+H accents, which made up about a third of the adult data. These two findings are in line with the prediction that the proportion of accents with steep pitch rises increases with age. However, unexpected in this respect is the finding that the five-year-olds produced more L+H* accents than the seven-year-olds. One possible explanation is that it is not rising accents in general that are difficult for young children to produce, but rather accents in which the lexically-stressed syllable is low and followed by a rise, as in L*+H accents. As explained in the background (section 1.3.3), previous studies that reported difficulties with rising accents (Local, 1980; Loeb & Allen, 1993; Snow, 1998) did not specify the type of accents; it is therefore possible that the contours that the children had to imitate happened to be ones containing L*+H accents.

Differences between adults and children emerged also on the phonetic level. Compared to adult speakers, seven-year-olds' accents had a smaller excursion, and children of both age groups produced accents with a slower speed of pitch change (slope) than the adults. Furthermore, they aligned pitch minima earlier than the adults, while the location of the pitch maxima was adult-like. These patterns suggest that children start to raise F0 earlier to reach the high target, compensating for the slower speed of pitch change. Slower speed of pitch change (i.e., shallower slope) in rising accents in child speech has been observed before (Loeb & Allen, 1993; Snow, 1998), albeit

3.7 General Discussion

only in sentence-final position. Researchers have tried to explain this phenomenon by the increased physiological effort in rising accents (cf. Lieberman's breath group theory, Lieberman, 1967; Snow, 1998). Indeed, it has been found that laryngeal and respiratory function in children for controlling loudness (sound pressure level) becomes adult-like only by age twelve to fourteen (Stathopoulos & Sapienza, 1997). Physiological constraints may have prevented the children from realizing pitch accents in a fully adult-like fashion. Alternatively, children may not have problems with controlling the speed of pitch change in general, but rather with the coordination of glottal activity and supralaryngeal articulatory motor control. This does not seem implausible, given that seven-year-olds' temporal coordination of the oral articulators themselves has been found to differ from that of adults (e.g., Cheng, Murdoch, & Goozée, 2007). Experiments comparing children's performance in consonant-vowel (CV) sequences and sustained vowels could resolve this issue. If speed of pitch change in CV sequences were slower than in sustained vowels, this could mean that there are problems with the coordination. If, however, speed of pitch change were comparable in CV sequences and sustained vowels, but generally lower than in adult speech, this would mean that children's pitch control is not yet adult-like.

The third result concerns the effect of contrast on topic marking. Contrast, here defined in terms of a change of subjecthood from the preceding sentence, did not affect the choice of pitch accent type in any of the three age groups. This is in accordance with Braun's (2005, 2006) findings of an equal distribution of L+H* and L*+H in both contrastive and non-contrastive contexts. However, unlike the speakers in Braun's study, the speakers in this study did not consistently align peaks later in contrastive contexts. There was also no effect of contrastiveness on pitch excursion or slope in any of the age groups, as would have been expected on the basis of Mehlhorn's (2001) findings. It should be kept in mind, though, that the criterion for contrast applied here is rather crude, and not fully comparable to the question-answer pairs in Mehlhorn's experiment or the carefully constructed contexts in Braun's study (which were in addition constructed to elicit double contrasts in both topic and focus constituents). Change of subjecthood is only one potential source for

3.7 General Discussion

contrast. A speaker can choose freely whether to express a given opposition (such as referent A – comment B vs. referent C – comment D) intonationally or not, an optionality that has already been pointed out by Bolinger (1972).

To summarize the findings of this study: first, the intonational patterns for topical referents in unscripted speech are richer than in scripted speech. Hence, working with unscripted data is essential for obtaining more reliable information about the relationship between information structure and intonation. Second, when children produce sentence-initial topical referents, they use the same pitch accents as adults, but in different distributions. It has been argued that this may be due to the articulatory difficulty of producing steep rising accents. Thus there appears to be no special way in which children – as compared to adults – mark topical referents. Finally, the phonetic realization of phrase-initial pitch accents is still under development until at least age seven. This may be because of increased physiological effort in producing rising pitch movements, or because children have not yet fully learned to coordinate pitch movements with the articulation of segments.

4 The marking of information status in narrative discourse⁴²

4.1 Introduction

This study investigates how German five- and seven-year-old children and adults use intonation to signal the information status of referents in narrative discourse. For the reader's convenience, the most important theoretical claims and empirical findings discussed in the background are summarized here again.

The standard assumption for languages like English and German has been that new referents are accented, whereas given referents are deaccented (Brown, 1983; Chafe, 1974; Prince, 1981). However, it has also been observed that given referents tend to be accented when they are re-introduced into the discourse at a later point (Brown, 1983; Hirschberg, 1993; Terken, 1984). This has led some scholars to argue that viewing the pragmatic given/new distinction as a dichotomy is too simplistic (e.g., Ariel, 1990; Chafe, 1974; Gundel et al., 1993). Chafe (1994) assumes that there is at least a third category between given and new, which he calls *accessible*. The precise nature of the intonational marking of accessible referents, however, is not agreed upon. Chafe hypothesizes that accessible information is treated in the same way as new information (Chafe, 1994, p. 75). Lambrecht (1994), on the other hand, assumes that speakers can choose whether to accent an accessible referent or not, depending on several other discourse factors. More recently, it has been suggested that the type of accent also plays a role in signaling a referent's information status, and that there may even be a distinct accent which is used to mark accessible referents. According to Pierrehumbert and Hirschberg (1990), H* pitch accents signal that a referent should be treated as new to the discourse

⁴² Parts of the research presented in this chapter have appeared in Herbst, L. E. (2007). German 5-year-olds' intonational marking of information status. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)* (pp. 1557-1560).

4.1 Introduction

(“added to the mutual belief space”, p. 290), whereas L* accents convey that the referent is already part of the listener’s discourse model. This hypothesis has received empirical support in eye-tracking experiments that have shown that the type of pitch accent listeners hear influences their interpretation of a referent’s information status even before the entire target word has been heard (Chen et al., 2007). For German, Baumann (2006) observed that in read speech, accessible referents are mainly realized with H+L*. Furthermore, it was shown that listeners have preferences for certain pitch accent types as realizations for different information statuses (Baumann & Grice, 2006; Baumann & Hadelich, 2003), including a preference for H+L* as a marker for certain types of accessibility. Based on these results, Baumann and Grice (2006) have put forward the following mapping of information status and accentuation type for German (repeated from Figure 6):

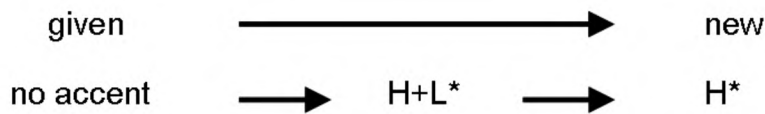


Figure 10: Proposed mapping between information status and accentuation type in German (adapted from Baumann & Grice 2006, p. 1655).

However, the level of detail with which information status marking in adult language has been looked at is not matched in acquisitional studies. Previous studies in this field suggest that young children’s information status–intonation mapping is similar to that of adults. On the perception side, Arnold (2008) has shown that four- to five-year-old children – like adults – have a bias to interpret deaccented nouns as referring to given (previously mentioned) referents. On the production side, Wonnacott and Watson’s (2008) experiment has provided first instrumental evidence that children use intonation to mark newness and givenness of referents. However, as already pointed out in the background, this study leaves open the question of how children mark information status in more complex discourse structures, and when discourse referents are not visually accessible for speaker and hearer. Furthermore, the study has not looked at the type of pitch accents the children used.

Finally, the authors did not compare the children's production with that of adults, or of older children. Thus the question remains open as to whether children use the same types of accents as adults do, and how their use of different accents may change with age.

This study aims at filling this gap. As outlined in chapter 2, a picture story telling task was used to elicit natural data. By analyzing more natural speech, the study also tests claims about specific accent types for information status marking in German that have so far only been supported by reading or perception experiments (Baumann, 2006; Baumann & Grice, 2006; Baumann & Hadelich, 2003). Unlike previous studies on child intonation in this area (Chen, 2007), the present study combines both phonological and phonetic analyses.

In the next section, the design of the materials and the elicitation procedure are described in more detail, and information about the data pre-processing and criteria for item selection are provided. In section 4.3, the prosodic annotation procedure is explained; it is shown that the intonation labels assigned are supported by quantitative acoustic evidence. In the phonological analysis, the types of pitch accents that the different age groups used are investigated. The fact that the present study is concerned with the use of intonation patterns in the context of real discourse has as the consequence that it is unavoidable to also look at the discourse-structuring function of intonation, which might interact with the way in which intonation is used to convey information about the information status of referents. More specifically, the present study will also look at the use of boundary tones. This will make it possible to also draw conclusions about how children deal with this dimension of intonation. Finally, there will be an investigation into children's phonetic realization of pitch accents and how it compares to that of adults.

4.2 *Method*

4.2.1 **Participants**

The participants were the same as described in chapter 3.

4.2 Method

4.2.2 Materials and design

Eight of the picture stories used to elicit the corpus (see chapter 2) were created to elicit the production of discourse referents with three distinct information statuses: new, given, and accessible. In the *new* condition, the target referent occurred for the first time. In the *given* condition, a referent was first introduced, and occurred again as target referent in the immediately following picture. In the *accessible* condition, a referent was introduced at the beginning of the story, and re-occurred after four (in one story) or five (in three stories) other intervening pictures. According to Baumann (2006), who follows Yule (1981) and Chafe (1994), a referent becomes semi-active if it has not been mentioned in the last two or three sentences. Presuming that each picture would be verbalized with at least one sentence, four to five pictures most likely suffice to lower the referent's activation. It has also been postulated that multiple referents in the speaker's mental discourse model compete for attention, lowering the activation of each one (Arnold & Griffin, 2007). Therefore, two additional referents – one animate and one inanimate – were introduced in the plot between the two occurrences of the target referent, which was expected to also lower the target referent's activation. Recall that the number of pictures in the story was varied between five and eight; this was achieved by adding pictures at non-critical parts of the story (e.g., following the target picture in the accessible stories).

Four different target referents were used, which all referred to animals that young children are likely to know. The target words were disyllabic and had a sonorant segmental make-up to facilitate pitch analysis: *Kamel* (/kɑ'me:l/ 'camel'), *Möwe* (/mø:və/ 'seagull'), *Biber* (/bi:bə/ 'beaver') and *Biene* (/bi:nə/ 'bee').

One of the aims of this study is to compare the intonational marking of information status of discourse referents in natural speech with that of the findings from the production and perception experiments in German (Baumann, 2006; Baumann & Hadelich, 2003). In these studies, the target referents were always in phrase-final position, because this is where the main accent of a sentence is typically found and where the intonational marking of a referent is

consequently “clearest in its form and most distinctive in its function” (Baumann, 2006, p. 95). The stimuli in the present investigation were therefore also created in such a way as to elicit referents in phrase-final position. The target pictures showed simple transitive events (e.g., hugging, drawing), in which the target referents had non-agentive semantic roles (e.g., recipient, beneficiary). This way the target referents were very similar to the target referents in the Baumann and Hadelich (2003) and Baumann and Grice (2006) studies, in which they were also always non-agentive.

Since the intonational phenomena of interest can be best observed on full noun phrases, several measures were taken to reduce the likelihood of pronoun use for the target referents. It has been observed that in narration, young children pursue a “thematic subject strategy”: They seek to identify the main protagonist of a story, which they then tend to encode with a pronominal expression; this happens less often with more peripheral characters (e.g., Bamberg, 1987; Karmiloff-Smith, 1979).⁴³ The stories were therefore constructed in such a way that the target referents played only a secondary role in the plot. In addition, the target referent and the agent shared the same grammatical gender so that pronoun use would lead to ambiguity, which is largely avoided, at least by adult speakers (e.g., Arnold & Griffin, 2007). Finally, the semantic role of the target referent switched from agentive to non-agentive in the given condition to discourage a “parallel role strategy”, whereby children sometimes use pronouns when co-referring expressions occur in the same

⁴³ Of course adults may also encode the main protagonist of a story with a pronoun. However, their pronoun usage is also governed by constraints of local cohesion: They tend to ensure the pronoun’s antecedent can be identified unambiguously. That is, if there were a switch in reference from one character to another, adults would typically encode this with a full lexical NP, irrespective of the character’s role in the story line. For example, in a story mainly about a baker, following sentences such as “The baker was on his way to the grocery store when he met the plumber. The plumber appeared to be drunk again”, adult speakers are more likely to subsequently switch reference to the baker again by using the expression “the baker” than using “he”. In contrast, children seem to consistently use pronouns to refer to the main character, even if this results in ambiguities.

4.2 Method

role in consecutive events (Hickmann, 1982). An example of a given story is provided in Figure 11.

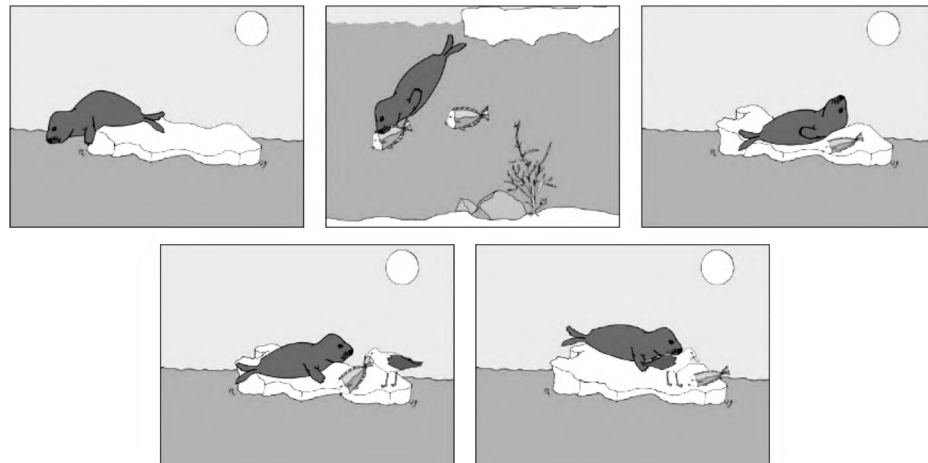


Figure 11: Example of a picture story (order from top left to bottom right) for eliciting a referent with given status. The target referent is the seagull introduced in picture 4, where it has an agentive role (stealing the fish). The seagull occurs again in picture 5 (the target picture), where it is an experiencer (bitten by the seal).

Each referent appeared once in each condition. The new and the accessible condition were elicited with the same story. This means that the introduction of the referent served as the target referent for the new condition and its re-introduction as target referent for the accessible condition. This was done to keep the experiment within reasonable length. An illustration of a new/accessible story is provided in Figure 12 below. Referents in the given condition were presented in a separate story. Four additional stories (three pictures each) from another experiment⁴⁴ served as fillers. In total, each participant told twelve stories. Four different lists were constructed by varying the order in which the stories appeared. Stories containing the same target

⁴⁴ This experiment is presented in chapter 5.

4.2 Method

referent (e.g., the new/accessible story containing *Kamel* and the given story containing *Kamel*) were always separated from each other by three other stories. Each time the story with the new and the accessible referent preceded the story with the given referent. Participants were randomly assigned to one of the four lists.

4.2 Method

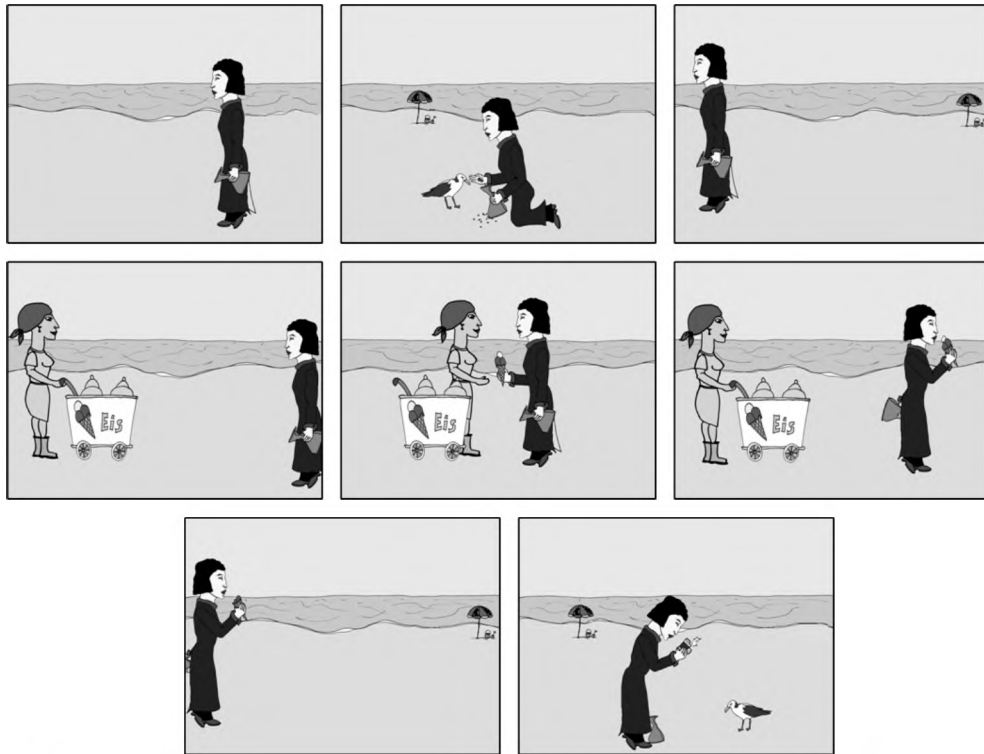


Figure 12: Example of a picture story (order from top left to bottom right) for eliciting a referent with new and with accessible status. The target referent is the seagull introduced in picture 2 (target picture for the new condition). The main protagonist (the woman in the dark dress) then interacts with another animate referent (ice cream seller) and performs an action on an inanimate referent (buying and eating an ice cream). The target referent is absent in these parts of the story. It re-appears in picture 8 (the target picture for the accessible condition). Note that in both target pictures the target referent fulfills a non-agentive role: It is an experiencer in picture 2 and a theme/experiencer in picture 8.

Children. As explained before in chapter 2, the children were tested individually by two female experimenters (Experimenter A and Experimenter B) in a quiet room at their kindergartens or schools. Each child participated in one

20-35min session. In case of the five-year-olds, the experimenters spent some time before the actual session playing with the children in their playgroups to familiarize the children with their presence. Experimenter A explained the procedure first to the children individually in the experimental room to make them feel comfortable with the situation. The children were informed about the story-telling game already described in chapter 2. After the child had signaled that she had understood the game, Experimenter B was invited to enter the room, and Experimenter A explained the procedure again to both. Children sat at a comfortable distance in front of the touch screen. The screen was positioned in such a way that neither Experimenter A nor Experimenter B could see it, and it was made clear to the children that they had a privileged view.

In each trial, Experimenter B asked a broad-focus question (e.g., *And what happens then?, Oh, and what comes now?* etc.) as a prompt before the picture intended to elicit the target referent. If the child's answer was unclear or ambiguous, Experimenter B indicated that she did not understand and asked the child again. This was repeated if necessary. If after two attempts the child still failed to provide a more informative answer, Experimenter B asked a more specific (narrow-focus) question (e.g., *Now, what is the seal biting?*), or gave the impression that she had misunderstood (e.g., *I see, the seagull is biting the seal*), which usually prompted the children to correct her. The purpose was to remind the children that Experimenter B did not see the pictures and to motivate them to provide more elaborate narrations next time, but these trials were later excluded from the analysis. Experimenter A took care not to mention any of the target referents. There was a break of 3-5 minutes after half of the trials.

4.2.3 Data pre-processing and selection

All referents were coded for their information status (new, given, accessible). Then, all utterances were annotated for whether they contained the intended target word (T), another full NP (O), a pronoun (P), or whether the referent was not mentioned (NM) at all. This last category concerned cases in which a speaker described a scene using gestures and deictic expressions (e.g., *Und*

4.2 Method

jetzt macht die so, ‘and now she is doing like that’) or generic descriptions (e.g., *Jetzt malt er*, ‘Now he is painting’). The breakdown for each of the four referents for the two age groups is given in Table 6 and Table 7.⁴⁵ These counts are irrespective of whether an item was excluded at a later stage, for example because of disfluencies or because the utterance was a response to a narrow focus question.

referent		possible	T	O	P	NM
Biene	N	86 ⁴⁶	45	30	8	4
	%	100	52.3	34.9	9.3	4.7
Biber	N	87	43	32	11	1
	%	100	50.0	37.2	12.8	1.2
Möwe	N	86	16	65	3	3
	%	100	18.6	75.6	3.5	3.5
Kamel	N	87	74	9	4	0
	%	100	86.0	10.5	4.7	0.0
TOTAL		346	178	136	26	8
(mean %)			(51.7)	(39.5)	(7.6)	(2.3)

Table 6: Number of possible items for the five-year-old group, actual (absolute and relative) frequency of target word uttered (T), other nouns (O), pronouns (P), and cases in which no NP was mentioned (NM), broken down by target referent.

⁴⁵ Note that these numbers are based on the children’s “final” productions following potential clarification questions by the listener. This means that it is possible that a child first produced a pronoun or failed to mention the referent, but then provided a “target” answer in the second attempt.

⁴⁶ The experiment with one child had to be aborted half through the session, so that there is data only for two (*Biber*, *Kamel*) of the four referents.

referent		possible	T	O	P	NM
Biene	N	81	50	20	5	3
	%	100	61.7	24.7	6.2	3.7
Biber	N	81	63	13	0	2
	%	100	77.8	16.0	0.0	2.5
Möwe	N	81	49	27	2	0
	%	100	60.5	33.3	2.5	0.0
Kamel	N	81	74	3	1	0
	%	100	91.4	3.7	1.2	0.0
TOTAL		324	236	63	8	5
(mean %)			(72.8)	(19.4)	(2.5)	(1.5)

Table 7: Number of possible items for the seven-year-old group, actual (absolute and relative) frequency of target word uttered (T), other nouns (O), pronouns (P), and cases in which no NP was mentioned (NM), broken down by target referent.

The tables show that the seven-year-olds produced more target words on average than the five-year-olds (72.8% vs. 51.7%). It is clear that the word *Möwe* was the most difficult for both age groups. In the five-year-old group, about $\frac{3}{4}$ of the occurrences of that referent were substituted with a different noun, which in most cases was the superordinate expression *Vogel* ('bird'). The seven-year-olds also produced fewer pronouns than the five-year-olds (2.5% vs. 7.6%), but pronoun use was generally low, as intended by the design. The five-year-olds used pronouns most often in the *Biber* immediate story.

4.2 Method

referent		possible	T	O	P	NM
Biene	N	78	58	16	2	2
	%	100	74.4	20.5	2.6	2.6
Biber	N	78	66	7	2	3
	%	100	84.6	9.0	2.6	3.8
Möwe	N	78	73	5	0	0
	%	100	93.6	6.4	0.0	0.0
Kamel	N	78	68	3	3	3
	%	100	87.2	3.8	3.8	3.8
TOTAL		312	265	31	7	8
(mean %)			(84.9)	(9.9)	(2.3)	(2.6)

Table 8: Number of possible items for the adult group, actual (absolute and relative) frequency of target word uttered (T), other nouns (O), pronouns (P), and cases in which no NP was mentioned (NM), broken down by target referent.

In the seven-year-old group, pronouns occurred slightly more often in the *Biene* immediate story, where some of the children described the scene with *Und jetzt haut die die* ('and now she's beating it'). Here it could be argued that the constellation of referents disambiguated the pronouns (recall that both agents have the same grammatical gender in German). It is clearly more likely for the human protagonist (the woman) to beat the bee than the other way round. The fact that the children picked up on this shows that they take the pragmatics of the situation into account. The adults produced the intended target word in about 85% of the cases. The most frequent substitutions of a target word by another noun occurred in the new/accessible story for the target referent *Biene*, in which many speakers used *Wespe* ('wasp') instead.

Since especially in the youngest age group often produced other nouns than the intended ones, the phonological analysis comprised both target and non-target words to arrive at an acceptable sample size. Pronominal forms (i.e., personal and demonstrative pronouns), however, were excluded. Four

items had to be excluded because of experimenter error.⁴⁷ The intonation of 25 items could not be analyzed because of creaky/breathy voice or laughter. Furthermore, utterances with disfluencies were discarded, if the disfluency could be classified as a re-start or a filled pause (*ähm* etc.) and if it occurred directly before or in the target word (e.g., *Und jetzt macht die ein Foto vom Kam- Kamel*), but not if the disfluency occurred at the beginning of the utterance (e.g., *Und.. äh...ja, dann malt er eine Biene*). If the disfluency was a hesitation (silent pause) which did not have a disruptive effect on the F0 contour if the pause was cut out, the utterance was included in the analysis, following a convention that has for instance been used in the Dutch ToDI System (Gussenhoven, 2005). On the basis of these criteria, 55 further items were excluded. I also excluded 17 elliptical utterances, that is, utterances that consisted only of the target word, 19 self-corrections (e.g., *Now she is taking a picture of the crow – no, the seagull*), and 18 utterances that had been produced as a response to a narrow focus question by the experimenter.

In addition, referents in the new condition had in 14 cases been realized with a definite article or a possessive pronoun (e.g., *Und jetzt setzt er sich auf sein Kamel*, 'and now he is sitting on his camel'). These were excluded, as it could not be ruled out that the speaker had perceived the referent as given or inferable from the preceding context. In 22 narrations of an accessible story, speakers did not produce the referent with the (expected) definite article. Of these, most came from the *Möwe* story. It seems that the speakers did not realize that the bird was the same they had introduced at the beginning of the story. Even though the seagull had been drawn with some characteristic features in order to make it unique, these were apparently not striking enough to make it obvious to all participants that the picture was about the same animal. Since in all these cases there was no co-reference between the referent that was first introduced and the referent that appeared later in the story, they were excluded from the analysis. I further excluded 16 cases in which speakers mentioned the target referent in the accessible condition before they were

⁴⁷ Here the experimenter had mentioned a target referent before the child had mentioned it.

4.2 Method

presented with the target picture (e.g., *Er geht wahrscheinlich zurück zu seinem Kamel*, 'He is probably going back to his camel'). These items were not analyzed. Finally, in 40 cases a different label was used for a referent in the accessible condition than the one with which the referent had first been introduced (e.g., *wasp* – *bee*), so they were excluded.⁴⁸

One of the aims of this study is to compare the intonational marking of information status of discourse referents in natural speech with findings from previous studies on German (Baumann, 2006; Baumann & Grice, 2006; Baumann & Hadelich, 2003). In these experiments, the authors looked at phrase-final discourse referents. As mentioned before, the stimuli in this study were therefore also designed to elicit referents in phrase-final position. In about half of all utterances the word denoting the target referent occurred phrase-finally. In these cases, the end of the phrase was marked by a clearly perceptible pause. However, in the other half of the utterances, the target word⁴⁹ was followed by one or several words. For some of these cases, it was difficult to decide whether the material following the target word was still part of the phrase in which the target word was embedded or not. For these cases, the decision was based on the results of a perception experiment, described in chapter 7. In this experiment, untrained listeners listened to the utterances and indicated on a written transcript where they perceived prosodic breaks. After sorting the utterances into final and non-final ones, I ana-

⁴⁸ The majority of these cases occurred in the *Biene* new-accessible story (see Figure 30 in Appendix A). Speakers often introduced the referent in a sentence like *He is painting a bee*, but later described the last scene as *Now he is cutting out his artwork*. Arguably, a referent like *artwork* is inferable from the painting action, and in this respect, it is also an accessible referent. However, these inferentially accessible referents are different from the other cases of (textually) accessible referents in that the noun is not "lexically given"; the word *artwork* has not been mentioned before. For this reason, these cases were not taken into the analysis.

⁴⁹ Note that target word here refers to the word denoting the target referent in the story, irrespective of whether the actual word was the intended one (e.g., *bee*) or a different noun (e.g., *wasp*).

lyzed the type of elements that followed the target word. The most frequently occurring types together with some examples are listed in Table 9.

type	Examples	English translation
modifier	Jetzt hat er den Kamel <u>wieder</u> . now has he the camel again [...] umarmt den Biber <u>aus Dankbarkeit</u> . hugs the beaver out gratitude	'Now he has the camel again.' '[...] hugs the beaver out of gratitude.'
non-finite verb-form	Die tut die Biene <u>kaputtmachen</u> . she does the bee destroy [...] fängt an, die Möwe <u>zu beißen</u> . begins the seagull bite	'She's destroying the bee.' 'She starts biting the seagull.'
verb particle	Der schneidet die Biene <u>aus</u> . he cuts the bee out Macht die die Biene <u>tot</u> . makes she the bee dead	'He's cutting the bee out.' '[She's] killing the bee.'
coordinated clause	Er nimmt sein Kamel und reitet. he takes his camel and rides [...] bedankt sich beim Biber <u>und umarmt ihn</u> . thanks himself with-the beaver and hugs him	'He takes his camel and rides.' 'Thanks the beaver and hugs him.'

Table 9: Examples of the most frequently occurring types of elements following the target word.

Table 10 shows the frequency of occurrence of the different types of post-target word elements. The totals indicate the number of items that were retained after the selection criteria described above were applied.

4.2 Method

element following target word	5 years		7 years		adults	
	N	%	N	%	N	%
∅ (phrase-final)	111	59.0	128	53.1	89	35.6
modifier	14	7.4	29	12.0	26	10.4
non-finite verb	15	8.0	10	4.1	26	10.4
verb particle	20	10.6	21	8.7	9	3.6
coordinated clause	6	3.2	10	4.1	30	12.0
other ⁵⁰	22	11.7	43	17.8	70	28.0
Total	188	100	241	100	250	100

Table 10: Frequency of elements following target word, broken down by age group.

In addition to the 328 phrase-final items, 101 non-final items were added to the data set. These were utterances in which the target word was followed by either a verb particle (e.g., *aus* ‘out’, *mit* ‘with’) or a non-finite verb form (e.g., *gemalt* ‘painted’, *machen* ‘made’). In these cases, the target referent is still the last referring expression in the sentence. This was the domain of analysis also in Baumann’s (2006) analysis of read-out newspaper texts, whose results were the point of departure for the above reported perception results (Baumann & Grice, 2006; Baumann & Hadelich, 2003). Thus the utterances analyzed here are very comparable to the ones analyzed by Baumann. Unlike Baumann, however, I also included these items if the final non-referential element carried an accent to increase the number of items. The final data set contained 429 utterances, 146 from the five-year-old group, 159 from the seven-year-old group, and 124 from the adult group.

⁵⁰ In the adult group, a large part of the “other” category consisted of combinations of several elements, such as modifier + non-finite verb.

4.3 Prosodic annotation and consistency check of accent labels

4.3 *Prosodic annotation and consistency check of accent labels*

The sentences containing the target referents were annotated and analyzed using the speech-analysis tool Praat (version 5.0.35, Boersma & Weenink, 1992-2008). Intonation of the words denoting the target referent, of words following those words, and of the subsequent phrase boundary was labeled following GToBI guidelines (Benzmüller & Grice, 1997; Grice et al., 2005). However, as pointed out before, intonation labeling is not without problems, as evidenced by the low agreement values obtained in Grice et al.'s (1996) study. Accurate pitch accent annotation is again crucial in the present study, and additional empirical support for the accent labels assigned is necessary. To this end, I adopted the procedure already described in chapter 3. From the defining features of the different pitch accent types given in the GToBI literature testable predictions were derived regarding phonetic differences between certain pairs of accents. Subsequently it was tested whether the accents that had been assigned different accent labels (e.g., L+H* and L*+H) do differ from each other in the predicted way. If this were the case, this would be evidence that the accents have been labeled consistently, and that the accents do constitute distinct categories.⁵¹

For ease of reading, these detailed analyses have not been included in the following text, but placed in Appendix B. The Appendix contains a list of predictions derived from the literature, details of the acoustic measurements and the statistical procedure, and a discussion of the results.

The results show that all potentially confusable accent pairs differed significantly from each other in line with the predictions. For example, it was shown that in accents that were labeled L+H*, the F0 minimum occurred before or at the beginning of the stressed syllable, whereas in accents that were labeled L*+H, the F0 minimum occurred clearly within the stressed syllable, as illustrated in Figure 13 below.

⁵¹ Note again that this refers only to the accent categories as postulated in the GToBI annotation system.

4.4 Results

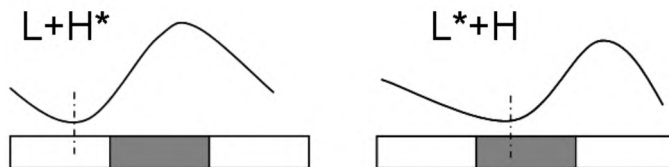


Figure 13: Schematic representation of the location of the F0 minima in the accents L+H* and L*+H. The grey-shaded area indicates the lexically stressed syllable.

However, it is possible that this does not hold for all three age groups alike, and that some of the effects are largely carried by only one or two age groups. In order to test this, I re-ran the same analyses for each age group separately. The details of the analyses can be found in Appendix C. Most of the effects reported in the overall analyses are also significant in the by-age group analyses. Of course the number of items in each accent category is considerably reduced, so that there is in some cases not sufficient statistical power for the effect to become significant; still, in these instances, the effects were numerically in the predicted direction.

The overall analyses and the individual age group analyses together provide substantial evidence that the accents have been labeled consistently and that the labels can be used for subsequent phonological analyses.

4.4 Results

Before turning to the main question of this study, the influence of information status and age on the use of different pitch accent types, I will first look at the distribution of boundary tones across age groups. Baumann (2006) and Baumann and Grice (2006) looked at low-ending utterances only. As it is unclear to which degree collapsing across different boundary tones might mask important patterns in the data, the type of boundary tone had to be taken into account in the analysis.

4.4.1 Boundary tones

Speakers used both high and low boundary tones at the end of their utterances. From looking at those items where the target referent was followed by

a boundary tone immediately (i.e., where there was no other pitch accent following), it becomes clear that the age groups differ in how often they produced utterances with a high or a low boundary tone (see Figure 14). Particularly striking is the distribution in the youngest age group, who produced nearly four times as many utterances ending in a low boundary tone than ending in a high boundary tone.

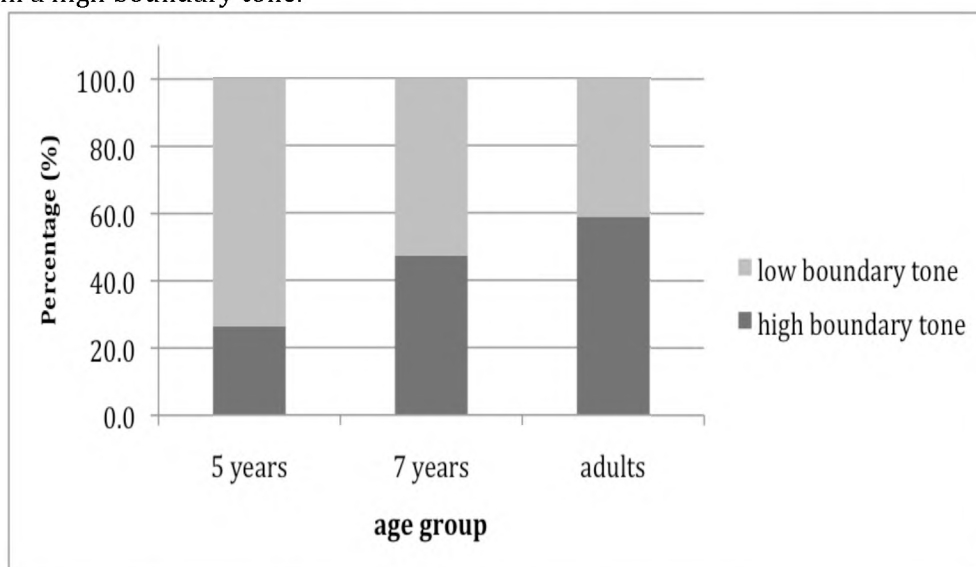


Figure 14: Distribution of high and low boundary tones by age group.

I tested these differences in boundary tone distribution statistically using a logit mixed-effect (LME) analysis (see Bates & Sarkar, 2007; Jaeger, 2008), with BOUNDARY (high/low) as the dependent variable and age (AGE) as the independent variable (fixed factor). As in all subsequent LME analyses in this chapter, subject (SUBJECT) and word (WORD) were initially included as random factors, but excluded if the variance explained by them was effectively zero. If the explained variance was not zero, likelihood ratio tests were used to test whether the random effect parameters were justified in the model (Agresti, 2002). All reported p-values were obtained by Laplace approximation.

4.4 Results

The analysis showed that there was a clear effect of AGE. Five-year-olds produced significantly fewer utterances with high boundary tones than both adults ($\beta = -1.75, z = -3.73, p < .001$) and seven-year-olds ($\beta = -1.04, z = -2.35, p < .05$). The seven-year-olds, on the other hand, did not differ from the adults ($p > .1$). The fact that both high and low boundary tones occur shows that it would not be a truthful representation of the data if only utterances with low boundary tones were analyzed. Furthermore, the unequal distribution across age groups indicates that there is a developmental shift between five and seven years in the way intonation is used to structure discourse. Before returning to this finding in the general discussion, let us first look at the main question of this study: Which pitch accent types did the speakers of the different age group use to signal information status?

4.4.2 Pitch accents

For pitch accent analysis, the categories were collapsed across boundary tones. If pitch accents were analyzed for low and high boundary tones separately, this would result in only few items in some categories, due to a strong association between the type of boundary tone and the type of pitch accent type: Low accents (L^*, L^*+H) frequently occurred with high boundary tones, and high accents with low boundary tones (see also Appendix B6). This is in itself an interesting finding, which we will return to in the general discussion. The suggested mapping between information status and pitch accent type is that new referents are signaled with a H^* accent, accessible referents are marked with the accent $H+L^*$, and given referents are deaccented. However, results show that – in addition to producing deaccented referents – speakers used seven different accent types. This finding already suggests that the range of intonation patterns found in natural speech production is more complex than what would be expected on the basis of the reading and perception studies presented above. The three most frequent accents and the alleged accessibility accent $H+L^*$ (see section 1.2) are illustrated in Figure 15.

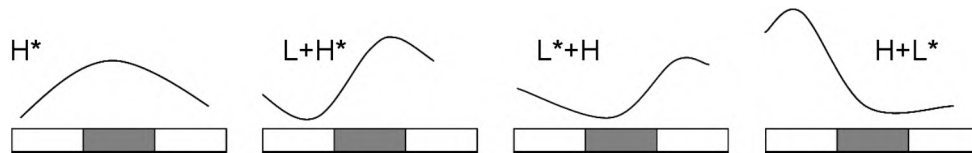


Figure 15: Shapes and GToBI labels of the three most frequent pitch accent types in the data, and the accent H+L*, followed by a low boundary tone. The grey-shaded box indicates the lexically stressed syllable.

Table 11 shows the distribution of the accents (including deaccentuation) by information status and age.

4.4 Results

		Status								
		New			accessible			given		
		Age			Age			Age		
		5	7	adults	5	7	adults	5	7	adults
		years	years		years	years		years	years	
deacc.	N	1	1	0	8	8	2	19	24	15
	%	1.8	1.8	.0	17.8	17.4	8.0	44.2	42.1	31.2
H*	N	15	16	11	13	20	3	8	16	12
	%	26.3	28.6	21.2	28.9	43.5	12.0	18.6	28.1	25.0
L+H*	N	14	15	9	8	4	10	3	3	2
	%	24.6	26.8	17.3	17.8	8.7	40.0	7.0	5.3	4.2
L*+H	N	5	13	15	5	6	6	1	3	4
	%	8.8	23.2	28.8	11.1	13.0	24.0	2.3	5.3	8.3
H+L*	N	0	1	3	1	0	2	0	1	5
	%	.0	1.8	5.8	2.2	.0	8.0	.0	1.8	10.4
H+!H*	N	10	4	7	7	4	0	6	5	6
	%	17.5	7.1	13.5	15.6	8.7	.0	14.0	8.8	12.5
!H*	N	7	1	2	2	1	0	6	3	2
	%	12.3	1.8	3.8	4.4	2.2	.0	14.0	5.3	4.2
L*	N	5	5	5	1	3	2	0	2	2
	%	8.8	8.9	9.6	2.2	6.5	8.0	.0	3.5	4.2
Total	N	57	56	52	45	46	25	43	57	48
	%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 11: Frequency distribution (absolute and relative) of pitch accent types by age and information status. The grey-shaded areas mark the most frequent realization for each age group in each condition.

For each of the eight accentuation types, I analyzed whether their frequency was affected by the referent's information status and the age of the speaker, using LME analyses. The variable accent type was recoded into binary variables (e.g., "H* yes/no"). In the models, this variable (henceforth ACCENT) was the dependent variable; information status (STATUS) and age (AGE) were the fixed factors (independent variables). The fitting procedure for random factors was the same as described previously. In a first step, a full model including both predictors (AGE and STATUS) and their interaction was specified. If the analysis did not bring about a significant interaction, the interaction was removed from the model. Predictors with a p-value larger than 0.1 were removed if this did not deteriorate the fit of the model (as estimated by a likelihood-ratio test). All reported p-values were obtained by Laplace approximation.

Note that mixed-effect models could not be used, however, where there were too few cases in one of the categories.⁵² In these cases, I used Generalized Linear Models (the GLM function in R) with only AGE and STATUS as factors instead.

Because the z-values from which the significance levels are derived in R are based on the Wald statistic, it is a problem for both LME and GLM if there are zero occurrences in one category (as discussed e.g., in Agresti, 1996, 2002; Menard, 1995). This is the case for the H+L* category in the seven-year-old group. To circumvent this problem, a z-test for two proportions was used instead, comparing the proportion of the accent type in each condition with its proportion in each of the other two. For this test, proportions need to be transformed into z-values (Fisher z-transformation) and then compared to each other (see Hinkle, Wiersma, & Jurs, 2003).

There were effects for six of the eight accentuation types, and only one significant interaction. Overall, there was a clear tendency for given referents

⁵² A rule of thumb for binomial models regarding the absolute upper limit for the number of parameters in the model is $\min(\text{outcome A, outcome B}) / 10$. Thus when there are only about twenty cases in one category, a model with three factors (two fixed, one random) would already be overfit.

4.4 Results

to be deaccented, while new referents were accented. Accessible referents assumed an intermediate rank regarding the proportion of not carrying an accent. Further effects of information status emerged for L*+H accents, which were more frequent for new and accessible referents than for given ones. There was only one interaction between age and information status, which showed that five- and seven-year-olds used L+H* accents often to mark new referents, whereas adults used them more often to mark accessible referents. With respect to effects of age, the analyses revealed three main differences. First, five-year-olds produced fewer L*+H accents than adults. Second, five-year-olds used !H* accents more often than adults. Third, both five- and seven-year-olds realized referents with H+L* accents to a much lesser extent than adults.

Below are the details of the analysis for the individual accentuation types.

Deaccented

STATUS was a significant predictor for the proportion of deaccented referents, but age was not. Given referents were deaccented significantly more often than both new referents ($\beta = -1.46$, $z = -4.46$, $p < .0001$) and accessible referents ($\beta = -4.24$, $z = -5.52$, $p < .0001$). Furthermore, accessible referents were deaccented more often than new referents ($\beta = 2.75$, $z = 1.59$, $p < .001$).

L*⁵³

Only STATUS turned out to be a significant predictor for the proportion of L* accents, but not AGE. L* accents were used significantly more often in the new condition than in the given condition ($\beta = 1.35, z = 1.99, p < .05$), but there was no difference between the new condition and the accessible condition ($\beta = -0.70, z = -1.19, p = .23$), or the accessible condition and the given condition ($\beta = 0.65, z = 0.83, p = .4$).

!H*

AGE was a significant predictor for the proportion of !H* accents, but STATUS was not. The five-year-olds used this accent type significantly more often than both the adult group ($\beta = -1.36, z = -2.35, p < .05$) and the seven-year-old group ($\beta = -1.32, z = -2.474, p < .05$). However, there was no difference between the seven-year-olds and the adults ($\beta = 0.04, z = 0.07, p = .94$).

⁵³ Although the results of the accent label consistency check (see Appendix B) suggest that the items labeled with the accents L*+H and L* were differentiated from each other by the position of the F0-minimum, Baumann (who was a reviewer for the article based on chapter 8) argues that this distinction cannot be made if the low target occurs in the ultimate or penultimate syllable. He assumes that all low accents preceding a high boundary tone should be labeled L*. If this is the case, then it is possible that observed effects for L*+H disappear if this category is merged with the L* category. To test this possibility, I repeated the analysis with a new merged L* category. The effects persisted: Both AGE and STATUS remained significant predictors for the proportion of L* accents. They were produced more often in the new condition than in the given condition ($\beta = 2.06, z = 4.79, p < .0001$), and more often in the accessible condition than in the given condition ($\beta = 1.56, z = 3.34, p < .001$). The new and the accessible condition, in contrast, did not differ significantly from each other in this respect ($\beta = -0.5, z = -1.44, p > .1$). Furthermore, five-year-olds used this accent significantly less often than the adults ($\beta = -1.50, z = -2.62, p < .01$), whereas no such difference was found between the seven-year-olds and the adults ($p > .1$). The two child age groups did also not differ from each other ($p > .1$).

4.4 Results

L*+H⁵⁴

Both STATUS and AGE turned out to be significant predictors for the proportion of L*+H accents. They occurred significantly more often in the new condition than in the given condition ($\beta = 1.80, z = 3.67, p < .001$), and they were also used significantly more often in the accessible condition than in the given condition ($\beta = 1.59, z = 3.00, p < .01$). No such difference was found between the new and the accessible condition ($\beta = -0.21, z = -0.56, p = .57$). The five-year-olds used these accents significantly less often than the adult group ($\beta = -1.44, z = -2.54, p < .05$), whereas the seven-year-olds did not differ from the adult group in this respect ($\beta = -0.70, z = -1.38, p = .16$). There also was no difference between the five- and the seven-year-olds ($\beta = 0.73, z = 1.29, p = .19$).

H+L*

H+L* accents were hardly produced by the child groups, precluding the use of LME or GLM. A LME model with only STATUS as fixed factor showed that it was not a significant predictor for the proportion of H+L* accents. To test for differences across age groups, z-tests were used. These tests confirmed that the five-year-olds realized referents significantly less often with this type of accent than adults did ($z = 3.17, p < .001$). The seven-year-olds used H+L* only twice, so they also differed from the adult group ($z = 2.94, p < .001$). In contrast, the five- and the seven-year-olds did not differ from each other ($z = 0.50, p = .30$).

L+H*

There was an interaction between AGE and STATUS regarding the proportion of L+H* accents. The five-year-olds used this accent more often in the new condition than in both the given condition ($\beta = 1.60, z = 2.21$) and the accessible condition ($\beta = 1.61, z = 2.21$). The same was true for the seven-year-olds (new vs. given: $\beta = 2.07, z = 2.91, p < .01$; new vs. accessible: $\beta = 1.47, z = 2.26, p$

⁵⁴ The effects persist also when the L*+H category is merged with the L* category (cf. footnote 53).

4.4 Results

< .05). Compared to the other two conditions, adults used L+H* accents significantly more often in the accessible condition (accessible vs. new: $\beta = 1.28$, $z = 2.18$, $p < .05$; accessible vs. given: $\beta = 2.96$, $z = 3.27$, $p < .01$). They used L+H* accents marginally significantly more often in the new than in the given condition ($\beta = 1.70$, $z = 1.92$, $p = .05$).

Both child groups realized L+H* accents more often in the new condition than the adults in the accessible condition (five-year-olds vs. adults: $\beta = 1.74$, $z = 2.20$, $p < .05$; seven-year-olds vs. adults: $\beta = 2.75$, $z = 3.15$, $p < .01$), but the adults also used L+H* accents significantly more often in the new condition than both the five-year-olds in the accessible condition ($\beta = 1.73$, $z = 2.19$, $p < .05$) and the seven-year-olds in the accessible condition ($\beta = 2.73$, $z = 3.14$, $p < .01$).

The main results of the analyses are schematically represented in Table 12, with significant differences highlighted in boldface.⁵⁵

⁵⁵ Very similar results are obtained if only the target words are analyzed (see Appendix D). Thus including non-target words in the present analysis increased the statistical power, but did not change the overall picture.

4.5 Within-accent comparison across age groups

accentuation type	differences by STATUS	differences by AGE
deaccented	new < accessible ⁵⁶ < given	5 years = 7 years = adults
L*+H	new > accessible = given	$ \begin{array}{ccc} 5 \text{ years} & < & \text{adults} \\ \diagdown & & \diagup \\ & = & \\ & & \diagdown \quad \diagup \\ & & 7 \text{ years} \end{array} $
H+L*	given = accessible = new	5 years = 7 years < adults
!H*	given = accessible = new	5 years > 7 years, adults
L*	$ \begin{array}{ccc} \text{new} & > & \text{given} \\ \diagdown & & \diagup \\ & = & \\ & & \diagdown \quad \diagup \\ & & \text{accessible} \end{array} $	5 years = 7 years = adults
L+H*	new > given, accessible for accessible > new > given for adults new > 5 years accessible = 7 years accessible 5 years new = 7 years new > adults accessible	5 years = 7 years adults

Table 12: Differences of pitch accent type distribution by information status and age. Statistically significant differences are highlighted in boldface.

4.5 Within-accent comparison across age groups

The GToBI labels that have been assigned to the accents in the corpus were supported by phonetic analyses (see Appendix B). It was shown that accent pairs that have previously been prone to confusion (e.g., H* and L+H*; L+H* and L*+H) were significantly different from each other along a number of acoustic parameters, as predicted from the GToBI literature. It was furthermore shown that these differences also held for each individual age group (see Appendix C). Accents that were labeled H* and L+H*, for instance, in the five-year-olds' data differ from each other as they do in the adults' data and in the seven-year-olds' data: L+H* accents have a higher excursion and a steeper slope. Thus speakers of all ages differentiated between the accent pairs that

⁵⁶ Interestingly, half of all deaccented referents in the accessible condition came from the *Biene* story. A possible explanation for this is presented in the general discussion.

4.5 Within-accent comparison across age groups

were contrasted with each other. However, this does not necessarily mean that the phonetic realization of one age group was the same as the other two. We have already seen in chapter 3 that at least for sentence-initial (i.e., pre-nuclear) accents on topical referents, there seem to be age-related differences. Compared to adult speakers, seven-year-olds' accents had a smaller excursion, and children of both age groups produced accents with a slower speed of pitch change (slope) than the adults. Furthermore, they aligned pitch minima earlier than the adults. In the following how the children's realization of the various pitch accent types in *phrase-final* position (i.e., nuclear accents) compares to that of adults is investigated. More specifically, whether the age groups differ in their produced pitch excursion and slope and (where applicable) their alignment of F0 minima and maxima is analyzed.

Unfortunately the sample sizes for the individual accent categories are rather small when they are broken down by age group, so that this investigation can only have an exploratory character. Still, a look at this data may reveal some interesting trends.

I analyzed those accent type + boundary tone combinations (nuclear contours) which occurred at least seven times in each age group:

- H* L-% (N = 40)
- L+H* L-% (N = 28)
- L*+H H-% (N = 48)
- H+!H* L-% (N = 34)

4.5.1 Analysis and results

For each contour type, I tested whether age (AGE) is a significant predictor for the phonetic parameter in question (i.e., F0 excursion, slope etc.). The models contained the parameter as the dependent variable and AGE as fixed factor. Subject was included as random factor. There were effects of AGE for three of the four contour types, which are reported below. Details of the analyses can be found in Appendix (E). No effect was found for the accent H+!H*.

4.5 Within-accent comparison across age groups

H* L-%

Adults produced H* accents that were followed by a low boundary tone with a larger excursion than both five-year-olds (3.1 st vs. 1.4 st, $p < .001$) and seven-year-olds (3.1 st vs. 2.0 st, $p < .05$). Surprisingly, five-year-olds produced these accents with a larger excursion than seven-year-olds (2.0 st vs. 1.4, $p < .05$). Adults realized H* accents also with a steeper slope compared to both the five-year-olds (17.6 st/sec vs. 6.9 st/sec, $p < .001$) and seven-year-olds (17.6 st vs. 7.1 st/sec, $p < .001$), while there was no significant difference between the two younger groups ($p = .73$). The differences in slope are illustrated in the box plot below (Figure 16).

L+H* L-%

Adults realized L+H* accents preceding a low boundary tone with a larger excursion than both child groups (6.9 st vs. 3.2 st in the five-year-olds, $p < .01$; 6.9 st vs. 4.1 st in the seven-year-olds, $p < .05$), but there was no difference between the five- and the seven-year-olds ($p = .39$). The same pattern is found for slope. The slope in adults' production of L+H* was about 22 st/sec steeper than in the five-year-olds' production (35.8 st/sec vs. 13.5 st/sec, $p < .01$) and about 19 st/sec steeper than in the seven-year-olds' production (35.8 st/sec vs. 16.5 st/sec, $p < .01$), whereas there was no difference between the two child groups ($p = .51$). The effects for slope are illustrated in Figure 16.

4.5 Within-accent comparison across age groups

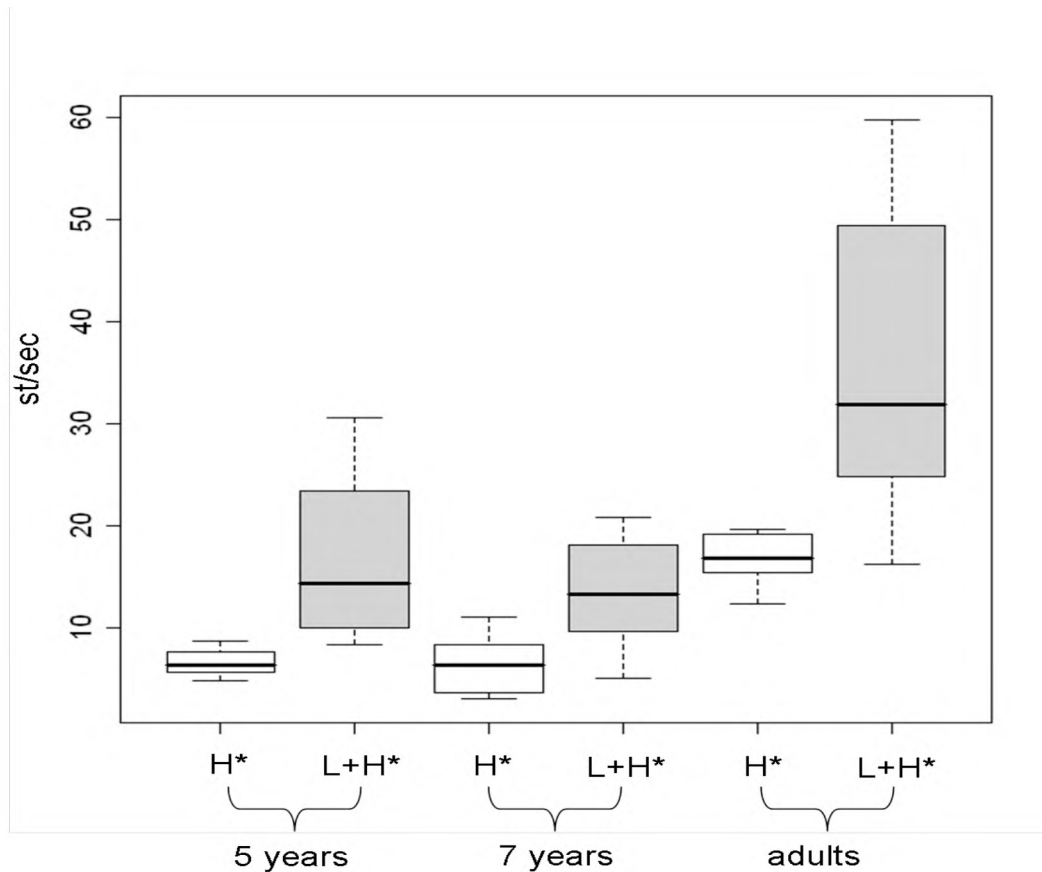


Figure 16: Box plot showing the distribution of values for the slope (in st/sec) in H* and L+H* accents (preceding low boundary tones) for each age group.

Furthermore, the children differed from the adults with regard to the alignment of the F0 minimum and the F0 maximum. Adults aligned the F0 minimum earlier than both five-year-olds (-8.9% w.r.t. the stressed syllable onset⁵⁷ vs. 6.8% into the stressed syllable, $p < .05$) and seven-year-olds (-8.9% vs. 6.4%, $p < .05$), as illustrated in Figure 17 below. There was no dif-

⁵⁷ A negative value indicates that the turning point precedes the onset of the stressed syllable.

4.5 Within-accent comparison across age groups

ference between the later two ($p = .91$). In addition, adults aligned the F0 maximum significantly later than the seven-year-olds (85.4% into the stressed syllable vs. 61.7%, $p < .05$), and marginally significantly later than the five-year-olds (85.4% vs. 68.8%, $p = .08$), but the two child groups did not differ from each other ($p = .54$).

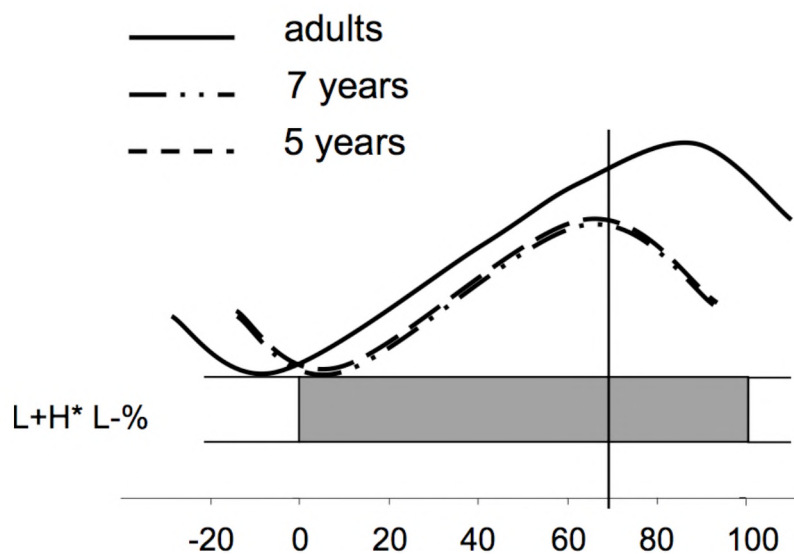


Figure 17: Graphical representation of the relative positions of the F0 minimum and F0 maximum in L+H* accents (followed by a low boundary tone), expressed as percentage of the stressed syllable duration (x-axis). The grey-shaded area marks the lexically stressed syllable. Note that slope is not to scale.

L*+H H-%

In L*+H accents (followed by a high boundary tone), no effect of AGE emerged for excursion, but there were again significant differences between adults and children with respect to the slope of the rise. Adults produced these accents with a steeper slope than both five-year-olds (34.7 st/sec vs. 22.34 st/sec, $p < .01$) and seven-year-olds (34.7 st/sec vs. 16.43, $p < .01$), but five-year-olds and seven-year-olds did not differ from each other ($p = .24$).

4.6 Discussion

This study investigated how German children and adults use intonation to mark the information status of discourse referents in natural speech. Two issues have been addressed: First, do adults use the types of pitch accents in natural speech production that have been postulated in the literature? Second, how do children use intonation for information status marking, and do they differ from adults?

Regarding the first issue, the results call into question a clear mapping between pitch accent type and information status, and more particularly the existence of a special accent for accessible referents. Speakers always accented new referents, which is consistent with claims in the literature (e.g., Chafe, 1974; Lambrecht, 1994) and previous production studies (e.g., Brown, 1983; Hirschberg, 1993; Terken & Hirschberg, 1994). However, no particular accent type emerged as the prototypical ‘newness accent’. The accent H*, which is assumed to be a marker for newness (Baumann & Hadelich, 2003; Pierrehumbert & Hirschberg, 1990) and the related L+H* were used by adult speakers in about 40% of the cases, but about 30% of new referents were realized with L*+H accents. This finding is at odds with the suggestion that L* accents signal givenness (Chen et al., 2007; Pierrehumbert & Hirschberg, 1990). In contrast to new referents, referents in the given condition were significantly more often deaccented, in accordance with previous reports (e.g., Brown, 1983). On the other hand, almost 70% of the given referents still received an accent, in line with previous findings that mere repetition of a referent does not necessarily lead to deaccentuation (e.g., Braun & Chen, submitted; Terken & Hirschberg, 1994). Interestingly, the proportion of L* accents was even lower in the given condition than in the new condition, which further questions the status of L* as conveying givenness. Even less evidence emerged for the existence of a special accessibility accent H+L*, as suggested by Baumann and Grice (2006). This accent type was used only once in the accessible condition, and was rare in the data overall. In this study, accessible referents patterned more with new referents than with given ones – in their preference to be accented, as claimed by Chafe (1994, p. 75), and

4.6 Discussion

partly in the type of pitch accent used (e.g., L*+H). Although adults used L+H* more often in the accessible condition than in the new condition, this accent was not used exclusively for accessible referents. The commonalities in the marking of new and accessible referents could be explained by Chafe's idea of activation cost mentioned in chapter 1. Both new (inactive) and accessible (semi-active) referents have a lower activation in the listener's consciousness: New referents, because their concept has not been activated at all, and accessible referents, because their concept is not activated anymore. In order to change an inactive or semi-active concept into an active one, the speaker needs to invest effort in terms of prosodic prominence (i.e., a pitch accent). The fact that accessible referents were not realized with a distinct accent and were sometimes also deaccented like many given referents supports Lambrecht's (1994) view that there is no clear phonological correlate for accessible referents, and that speakers decide whether to treat accessible referents like given or like new referents.

In the light of these results, the hypothesis that particular accent types are used to signal distinct information states seems difficult to maintain, at least with respect to natural speech production. Rather, the findings back the basic assumption that accentuation of new referents is obligatory, whereas the realization of given (and accessible) referents is less constrained. This view has been expressed by several researchers, albeit from a perception perspective: The absence of accentuation always signals givenness, whereas the presence of an accent does not constrain the interpretation in such a way (e.g., Birch & Clifton, 1995; Lambrecht, 1994; Terken & Nootboom, 1987).

Why then did other studies find that pitch accent type plays such an important role (Baumann, 2006; Baumann & Grice, 2006; Chen et al., 2007)? What seems crucial here is to make two distinctions: first, the distinction between speech perception under laboratory conditions on the one hand, and speech production on the other, and second the distinction between read speech and spontaneous speech.

The visual world paradigm (as used by Chen et al., 2007; Dahan et al., 2002) represents a restricted visual and linguistic environment. In such a setting, listeners need to quickly come up with an interpretation as to which is

the intended referent, because the task requires them to figure out where to move the computer mouse next (*Now move the CANDle above the square*). Perception studies like this are important, because they can tell us something about the possible meanings or functions that listeners associate with particular accents or contours (e.g., marking a referent as new or accessible). However, this does not necessarily mean that accents are exclusively associated with those meanings; these interpretations may only prevail in this particular setting. It is also important to recognize that the results obtained in this kind of experiment only indicate a *relative* preference to interpret a given accent in a particular way; it is never the case that all listeners always look at the same referent when hearing a particular accent. For example, in Chen et al.'s (2007) experiment, the highest number of mean fixation proportions to the new referent upon hearing H* was less than 0.3. In other words, although listeners may have interpreted H* as signaling newness more often than they did with other accent types, in more than 70% of the time when hearing H* they did *not* look at the new referent. This fact shows that even in this setting, there is no one-to-one mapping between intonation and information status.

A similar explanation could account for the differences between the present findings and the pitch accent type preferences obtained in offline judgment tasks (as employed by Baumann & Grice, 2006; Baumann & Hadelich, 2003). If different pitch accent types *can* convey different information states – and this is what the results from eye-tracking experiments like Chen et al.'s (2007) study suggest –, listeners will prefer one type to the other. However, this probably only holds as long as listeners attach importance to “canonical” information status marking in their judgments, and this is likely to depend on the context in which listeners perceive speech. The participants in Baumann and Hadelich's (2003) and Baumann and Grice' (2006) experiments listened to read-out stimuli. Previous research has shown that listeners are sensitive to “formal intonation” as a cue to identify speech as read-out rather than spontaneous (Blaauw, 1994; Laan, 1997). It is possible that clear information status marking is an important feature of intonation in read-out speech, and as such listeners would penalize intonation deviating from this in their ratings of read-out texts. But it may not be as relevant in other contexts, in which for

4.6 Discussion

instance paralinguistic functions of intonation play a larger role. It would be interesting to see if listeners judge the use of particular accent types such as H+L* differently when they have to rate unscripted speech.

If speaking mode (i.e., read vs. spontaneous) were to have an impact on the intonational marking of information status, this could also explain the discrepancies between the present study and Baumann's (2006) analysis of read-out newspaper texts. H+L* may be an appropriate marker for accessible referents in read speech, but may be used much less, or used in a different function (as discussed below) in spontaneous speech. Thus whether we find strong correlations between information status and pitch accent type may not only depend on modality (perception vs. production), but also on speaking mode.

The adults' narrations also differed from scripted speech in that they featured many high boundary tones, often in combination with L*+H accents (see Appendix B2). This rising pitch at the end of a clause typically indicates that there is "something more to come", and is commonly referred to as "continuation intonation" (e.g., Cruttenden, 1997; Schegloff, 1996). By using continuation intonation, the speakers signaled that the utterance was part of a larger structure, namely the story they were telling. This fact illustrates again the multifunctionality of intonation; it can be used to signal information status, but it is among others also a marker of discourse structure. In this connection it is also interesting that there was a high correlation between the type of boundary tone and the type of pitch accent, which has been noted before (Dainora, 2002b). High boundary tones occurred mostly with low pitch accents, low boundary tones occurred with high pitch accents. It seems that speakers have a preference to combine contrasting pitch accents and boundary tones. If this were the case, and if intonational meaning was truly compositional, one would like to know whether speakers first decide on the type of boundary tone they want to use, and then select the contrasting pitch accent accordingly, or vice versa. This rationale could perhaps explain why there were so many low pitch accents also for new referents – their use may have

followed from the speakers' decision to use high boundary tones to indicate continuity.⁵⁸

In view of this complexity, the findings regarding the second question of this study are remarkable. Children as young as five years old have learned to use intonation to mark information status, and they do it in an adult-like fashion. Children signal newness by accenting, and know that givenness (previous mention) licenses deaccentuation. This is in line with what Wonnacott and Watson (2008) reported for simple sentence pairs, but the results of the present study establish that this also extends to more complex discourse structures such as narrations. What is more, the results provide evidence that within such structures, children do not operate on a simple rule to treat every referent that has been mentioned as given. They realized accessible referents for the most part in the same way as new referents, just like the adult speakers. The differential treatment of given and accessible referents demonstrates that the children were sensitive to recency of mention, and used intonation to indicate this. The only difference between the children and the adults in information status marking concerns the use of L+H*. While the children used this accent more for new referents, adults used it more often as a marker for accessible referents. Given that L+H* is a perceptually more prominent accent than H*, it is possible that the adults wanted to draw the listener's attention more to the accessible referents than the children. Recall that the adults were instructed to tell the stories in a way that would enable the listener to retell them. The adults may have used L+H* to a greater extent to "remind" the listener of the reappearing referent – to invest more activation cost, in Chafe's terms.

⁵⁸ However, as mentioned in chapter 1, it is far from clear if intonational meaning is indeed compositional; in other words, we do not actually know whether pitch accent and boundary tone are functionally independent from each other. If they were not, then it would obviously not make much sense to talk about an order of selection. Rather, we would then have to assume that speakers choose an overall contour to convey a particular meaning.

4.6 Discussion

Another interesting finding concerning accessible referents is the fact that half of all deaccented accessible referents came from the *Biene* story – and this proportion was constant across all three age groups. What may have been special about this story to prompt relatively more deaccented realizations than the other stories? In this story (see Figure 30 in Appendix A), an artist paints a bee, then leaves the room for a brief break in the kitchen, and finally returns to his painting and cuts out the bee. What is different compared to all other stories is that everything takes place inside. One could speculate that in the other stories events were less predictable, as the main protagonist was moving about in open space. As a consequence, the reoccurrence of the target referent may have been more “newsworthy”, therefore increasing the speakers’ wish to “reactivate” its representation. Conversely, the fact that the artist returns to his painting may have been more predictable, leading in turn to more frequent deaccented referents. Another difference is that the bee is not really an animate referent in the same way as the other referents are animate. The picture depicts a referent that is typically animate (unlike a table or a hat), but it is still a painting. This meant that it could not have left the house, which makes its persistent presence more likely. If this reasoning was correct, it would indicate that even the youngest children register subtle context effects of this kind, and let predictability influence their intonational realization of accessible discourse referents.

Importantly, the results show that five-year-old children have already acquired a sizeable inventory of different pitch accent types. While the Wonnacott and Watson (2008) experiment provided global acoustic evidence that children manipulated pitch in nominal reference production, the present study reveals that this pitch manipulation is expressed in a number of different intonation patterns, similar to the adult inventory.

This early intonational competence in information status marking is in contrast with the difficulty children of the same age have been found to experience in the morphosyntactic domain, where they tend to use definite articles and pronouns inappropriately to introduce new discourse referents (e.g., Kail & Hickmann, 1992). However, as mentioned in the background, this difficulty may not stem from an inability to take the listener’s viewpoint into account.

This explanation would be difficult to reconcile with findings that even infants have the social-cognitive skills to know what is new for others (e.g., Tomasello & Haberl, 2003). It is more likely that the problems are due to the fact that children are only gradually becoming aware of the various functions that linguistic forms can have. It has been found that while the use of definite/indefinite forms to mark information status seems to emerge later, children aged three use these forms differentially to encode the distinction between specific and non-specific reference (e.g., Brown, 1973; Maratsos, 1974, 1976). The present results suggest that unlike with morphosyntax, five-year-old children have already learned the function of intonation to mark information status.

However, the data also revealed some age-related differences in the intonation repertoire, and I propose that these differences can be accommodated within a functional approach as well. Firstly, the youngest age group did not use phrase-final rising intonation (L^*+H/L^* accents together with high boundary tones) as much as the seven-year-olds and the adults. Several researchers have suggested that high pitch at the end of the utterance is also a “turn-keeping” device (e.g., Caspers, 2003; Couper-Kuhlen & Selting, 1996), that is, a signal that the speaker is about to say more. The fact that the five-year-olds did not produce this pattern as often as the older speakers may be an indication that children learn this discourse-structuring or interactional function of intonation only later. This is in line with a study by Potamianos and Narayanan (1998) who found that compared to older speakers (eleven- to fourteen-year-olds), even eight- to ten-year-old children produce fewer filled pauses in dialogue, which are an important device to signal delays in responding, thinking etc. in adult discourse (Clark & Fox Tree, 2002; Levelt, 1983; Schegloff, Jefferson, & Sacks, 1977). Potamianos and Narayanan’s finding is corroborative evidence that the acquisition of certain interactional devices (including continuation intonation) may be a late development.⁵⁹ It is also possible that

⁵⁹ Incidentally, my impression during testing was also that five-year-olds often did not signal when they were thinking or searching for words, which sometimes resulted in cross-talk between child and experimenter. The seven-year-olds, on the other hand,

4.6 Discussion

five-year-old children do use continuation intonation, but only in contexts in which they already know what the global structure of their contribution/discourse is. In the present setting, the stories were presented picture-by-picture, which could have discouraged them from indicating they were about to say more, since they did not know what would happen next. Looking at children's intonation in stories they are familiar with could test this hypothesis. These could either be achieved by giving them the chance to work through the stories beforehand, or let them narrate personal stories.

Alternatively, the relative scarcity of final rising contours – in particular $L^*+H\ H-\%$ – in the youngest age group may have physiological reasons. Imitation studies have found that four- to five-year-old children have difficulty producing rising patterns (e.g., Loeb & Allen, 1993; Snow, 1998), and have suggested that this is due to increased physiological effort. We have also seen in the preceding chapter that low rising contours of this kind were rare in the

appeared to produce signals such as “ähm” ‘uhm’ much more often. In order to see whether this impression could be backed up by the data, I have looked at the number of utterances that had been transcribed as “mmh”, “äh” and “ähm” in the narrations of the three age groups (excluding utterances that were clearly intended as feedback to the interlocutor's talk in the sense of “I understand”). Although this was only a cursory investigation, I found that there were differences across age groups: In the corpus of the five-year-olds, I counted 324 such instances, compared to 613 in the seven-year-old group. Interestingly, the number of these markers dropped again in the adults, who produced 327. The decreasing trend in adults seems at first surprising. However, it was also clear that compared to the children, the adults were more fluent and had much less difficulty with verbalizing the events on the pictures as they were going along, so that there were fewer delays in speech production that needed to be signalled. It is also possible that the adults perceived the situation as more formal (after all, they were aware that their narrations were being recorded), and therefore made an effort to produce fewer disfluencies. Potamianos and Narayanan unfortunately do not provide information about the number of filled pauses in their adult group, so it is unclear whether they observed a similar decrease.

five-year-olds' topic marking. However, the current data set cannot decide between these two alternative explanations.

Another interesting difference is that five-year-olds used the downstepped accent !H* significantly more often than the other two age groups. The phenomenon of downstep has been the topic of a number of investigations, but the majority of these are concerned with questions of speech production planning and phonetic realization (e.g., Grabe, 1998; Liberman & Pierrehumbert, 1984; van den Berg, Gussenhoven, & Rietveld, 1992). The function and meaning of downstep, on the other hand, has not been discussed much. Ladd (1996) states that "downstepping adds a nuance like finality or completeness, but does not otherwise seem to affect the meaning of the contour" (p. 76). If we follow this interpretation, the five-year-olds' increased use of !H* – or rather: the other age groups' decreased use of it – fits in nicely with the pattern found for the boundary tones and rising contours in general. Rising contours (which were produced more often by the older speakers) indicate continuity. Finality and continuity are opposites and cannot be signaled at the same time. If the older speakers were more often stressing the continuative aspect of their narrations, it follows from it that they would not use intonation indicating finality as much; the five-year-olds' patterns would be the mirror image of this.

The third difference between the child and the adult speakers was the frequency of H+L*, which occurred only once in the child data. Although it was also relatively scarce in the adult data, this difference between age groups was significant. I suggest that there are two possible explanations for this. First, adult speakers may use this accent type in spontaneous speech primarily to communicate a particular attitude (rather than a particular information status). As speculated above, it is conceivable that pitch accent types fulfill different functions in different contexts. Perception experiments have shown that H+L* has a "matter-of-fact" flavor to it and conveys an impression of "knowingness" (Dombrowski, 2003; Kohler, 1991). Given the simplicity of the story plots, it is not implausible that some adult speakers used this accent type to indicate that the event they were describing was to be expected, or obvi-

4.6 Discussion

ous.⁶⁰ It is possible that some paralinguistic uses of intonation to convey certain attitudinal meanings may be another aspect that is acquired only later. Findings that even seven- to nine-year-old children are unsuccessful in using prosody to understand irony, for example, point in this direction (Filippova & Astington, 2008). However, it is also possible that children do use intonation to signal attitudes, but that the materials and the discourse situation were not conducive to eliciting such patterns. Investigations of intonation production under different conditions (e.g., having children communicate with peers instead of adults) could shed more light on this issue.

An alternative yet related hypothesis is that H+L* accents are part of a certain register, such as “narrating”, which some of the adults made use of in this task, but children may not yet have acquired. In fact, one could imagine that precisely *because* H+L* does convey knowingness on behalf of the speaker, it is also employed in narrations, since in telling a story, the speaker often takes the position of an omniscient narrator.⁶¹ However, in view of the low numbers, the explanations offered here could only be speculative. More research is needed to determine the function of H+L* accents in different discourse situations, and, if it is register-specific, at which age children come to use this.

In the last part of the discussion I turn to phonetic aspects of intonation. It was also shown that the intonation labels could be supported by instrumental evidence: Accents that were labeled differently differed from each other also along acoustic parameters like pitch excursion in the way it would be pre-

⁶⁰ Note that H+L* occurred most often in the given condition (cf. Table 11) in which the main protagonist of the story performed an action on the target referent, as a consequence of the target referent’s preceding action. For example, in the given story presented before (see Figure 11) the seal is biting the seagull after the seagull had stolen a fish from the seal. It is conceivable that adults used intonation to mark this consequence as being “obvious”.

⁶¹ Interestingly, we also find allusions to H+L* as having a “soothing” note (Baumann & Grice, 2006; Grice & Baumann, 2002). It does not seem far-fetched to make a link between being knowledgeable and being able to comfort someone (“I know that everything will be alright”).

dicted by the GToBI literature. It was furthermore shown that the differences between pitch accent types held for each age group individually as well. However, the analyses revealed also that in three of the four nuclear contours investigated, the child groups showed divergence from the adults with respect to the phonetic realization. Generally, children produced accents with smaller excursions and shallower slopes than adults. In L+H* accents, there was furthermore a difference in alignment. Adults realized the pitch minimum earlier and the pitch maximum later than the children. Together with the larger excursion and the steeper slope produced by the adults, this points at a more pronounced realization of this accent type in adult speakers, or an “under-shoot” in the child speakers. These findings indicate that although the children have acquired most accents of the adult pitch accent inventory, and use these accents aptly for the marking of information-structural categories, their actual realization of those accents may still be slightly different from that of adults.

No differences in excursion or slope were found for the accent H+!H*. Recall that in H+!H*, the excursion and the slope were measured on the falling movement (namely from the preaccentual peak to the downstepped peak accent), whereas in the other three accent types, these measures were taken on the rising part of the accent. The finding that children have more difficulty with rising contours compared to falling contours has been reported before in the literature (e.g., Loeb & Allen, 1993; Patel & Grigos, 2006; Snow, 1998), where it has been attributed to the increased physiological effort involved in producing a rising pitch movement. The present results seem to indicate that this may still hold for seven-year-old children, although the sample sizes clearly limit the conclusions. On the other hand, the findings are in accordance with those obtained in chapter 3 for accents on topical referents, where it was also found that the children’s sentence-initial accents were produced with a smaller excursion and a slower speech of pitch change (slope). Evidence seems to converge that some physiological constraints may prevent children from realizing pitch accents in a fully adult-like way. Future research is warranted on how the development of the various subsystems (articulatory, laryngeal, respiratory) involved in speech production impacts children’s acquisition of adult-like pitch accent realization, and to what extent the difficulty

4.6 Discussion

associated with particular accent patterns might influence how often these accents are used.

To summarize the findings of this study: First, there is no one-to-one mapping between information status and pitch accent type in natural speech. Given referents have a tendency to be deaccented, but need not be. New and accessible referents are preferably accented, but the type of pitch accent is variable, and may be influenced by other pragmatic considerations such as the signaling of continuation. Second, five- and seven-year-old children do not differ from adults in the way they use intonation to mark information status in natural discourse. They tend to indicate givenness by lack of pitch accent, while realizing new referents with an accent and re-accenting accessible referents, using by and large the same set of pitch accents that adults use. Third, children seem to differ from adults in the use of other pragmatic and paralinguistic functions of intonation. Continuation intonation is less common in five-year-olds, and both five- and seven-year-olds do not seem to use intonation to signal attitudinal meanings or register as adults do. Finally, although children use by and large the same set of pitch accents, they still appear to have some way to go until the phonetic realization of these accents becomes fully adult-like.

5 The marking of identifiability and activation

5.1 Introduction

In sections 1.1 and 1.2, the fact that most theories of information structure distinguish between two basic types of givenness: identifiability and activation was discussed. However, the foregoing investigation of information status marking did not make this distinction. In fact, all new referents in the study were unidentifiable referents, that is, neither speaker nor addressee knew them before. The results showed that children seem to be as aware as adults of the necessity to mark the newness of this type of referents. But what about the marking of *discourse-new*, but identifiable referents? It is conceivable that this distinction is harder to grasp for children than for adults, and it is thus an interesting question whether children and adults signal new, but identifiable referents in the same way. This section presents an investigation of this question. For convenience, the following paragraphs first recap relevant parts of the background and terminology.

In chapter 1, we defined an identifiable referent as a referent of which the speaker assumes that the addressee already has a mental representation, or file-card. The representation may have been created in the course of the conversation by virtue of proper introduction (*I bought an interesting book yesterday*), in which case the referent could be said to be part of the local common ground between the interlocutors. But speaker and hearer may also already have a mental representation of the referent because it is part of their *communal common ground* (Clark & Marshall, 1981), that is, it is known by all members of the cultural community that speaker and hearer belong to (*The pope has had a stroke*).

Identifiability of a referent is typically signaled via morphosyntax, and especially by the use of definite determiners. According to Lambrecht (1994), an identifiable referent can have one of the three activation states, inactive, semi-active or active (see Figure 18, for convenience repeated here from Figure 3). For identifiable but inactive referents, Prince (1981) suggested the term *un-*

5.1 Introduction

used, for referents that are new for both speaker and hearer the term *brand-new*. In terms of activation, brand-new and unused referents should be equally inactive – brand-new referents because there is not yet a representation that can be active, and unused referents because the representation has not been activated in the course of the conversation. Consequently, when brought into the discourse, both types of referents are expected to be marked intonationally with a pitch accent.

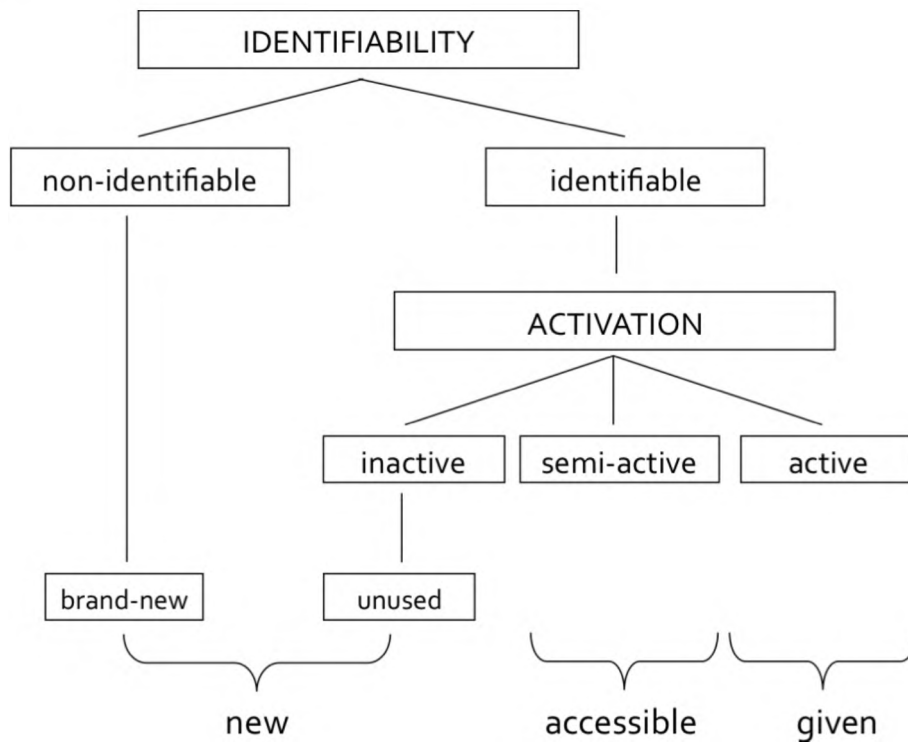


Figure 18: Taxonomy of givenness relations (based on Baumann, 2006, p. 68, following Prince, 1981 and Lambrecht, 1994).

The previous results have already shown that adults and children mark brand-new referents intonationally, preferably with (L+)H* accents. The present

5.2 Participants and materials

study tested whether children would mark unused referents in the same way, by having participants describe picture sequences which featured two referents that are well-known to all members of the “German cultural community”, as it were: *der Sandmann* (‘the sandman’) and *der Weihnachtsmann* (‘Santa Claus’). Hoping that the increased commercialization of Santa Claus and related incarnations has not (yet) convinced the majority of children that there is more than one instance of *Weihnachtsmann*, both referents should be taken to be uniquely identifiable. Could the fact that these referents are culturally given (and also of special importance to children) make their representations more activated than those of brand-new referents?

5.2 *Participants and materials*

The participants were the same as described in chapter 3. In the corpus already introduced in chapter 2 and described in more detail in chapter 4, two stories were created in the *new* condition that contained referents typically produced with a definite article, because they are uniquely identifiable: *der Sandmann* (‘the sandman’) and *der Weihnachtsmann* (‘Santa Claus’). The two stories were made up of three pictures each. As in the other stories, the target referents were minor characters with non-agentive semantic roles (e.g., theme), aimed at eliciting their realization in phrase-final position. An example is given in Figure 19 below.

5.3 Procedure, item selection and annotation

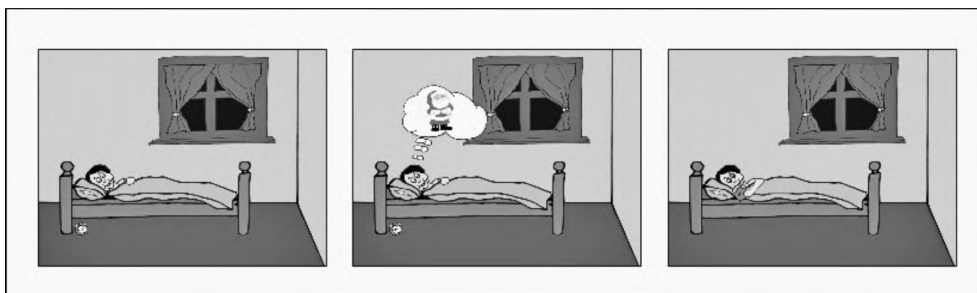


Figure 19: Story featuring a referent that is typically viewed as uniquely identifiable and produced with a definite article (*der Weihnachtsmann*, 'Santa Claus').

5.3 Procedure, item selection and annotation

The stories were presented along with the other stories described in chapter 4. The item selection procedure was the same as described in section 4.2.3. In particular, an item was excluded if the speaker did not produce the expected determiner type (e.g., *Er träumt von einem Weihnachtsmann*, 'He's dreaming of a Santa Claus'). The final data set contained 69 items, of which 22 came from the five-year-olds, 20 from the seven-year-olds and 27 from the adult group. Pitch accent type of the target words and the following boundary tone were again labeled following GToBI guidelines.

5.4 Results and discussion

Table 13 gives the distribution of the different pitch accent types that the speakers used for the identifiable but inactive (i.e., unused) referents. The results are presented together with the results for the new condition of the regular nouns in chapter 4.

5.4 Results and discussion

			Age					
			5 years		7 years		adults	
			type		type		type	
			unused	brand-new	unused	brand-new	unused	brand-new
Accent deacc.	N	0	1	0	1	0	0	
	%	.0	1.8	.0	1.8	.0	.0	
H*	N	8	15	3	16	3	11	
	%	36.4	26.3	15.8	28.6	11.1	21.2	
L+H*	N	6	14	9	15	4	9	
	%	27.3	24.6	47.4	26.8	14.8	17.3	
L*+H	N	3	5	3	13	10	15	
	%	13.6	8.8	15.8	23.2	37.0	28.8	
H+L*	N	1	0	0	1	4	3	
	%	4.5	.0	.0	1.8	14.8	5.8	
H+!H*	N	1	10	1	4	2	7	
	%	4.5	17.5	5.3	7.1	7.4	13.5	
!H*	N	2	7	1	1	2	2	
	%	9.1	12.3	5.3	1.8	7.4	3.8	
L*	N	1	5	2	5	2	5	
	%	4.5	8.8	10.5	8.9	7.4	9.6	
Total	N	22	57	19	56	27	52	
	%	100	100	100	100	100	100	

Table 13: Frequency distribution (absolute and relative) of pitch accent types by age and newness type. Unused referents are the target referents *der Weihnachtsmann* and *der Sandmann*, brand-new referents are the four regular nouns from chapter 4. The grey-shaded areas mark the most frequent realization for each age group and newness type.

In order to find out whether the frequency of a particular accent type was influenced by what we may call “newness type”, or by the age of the speaker, I

5.4 Results and discussion

performed binomial analyses using logit mixed-effect models (see 4.4.2), with ACCENT (e.g., “H* yes/no”) as the dependent variable, and AGE and (newness) TYPE as predictors (fixed factors). If a given accent type occurred more often with unused referents or with brand-new referents, there should be an effect of TYPE. If one of the age groups used a given pitch accent type more often in general, we would expect an effect of AGE. If one age group used a particular accent type more often with one of the two newness types, a significant interaction of these two factors should be observed. The fitting procedure was the same as described in section 4.4.2: In a first step, a full model including both predictors and their interaction was specified. The interaction was removed from the model if it was not significant. Predictors with a p-value larger than 0.1 were excluded if this did not deteriorate the fit of the model (as estimated by a likelihood-ratio test). Subject (SUBJECT) and word (WORD) were initially included as random factors. Likelihood-ratio tests were used again to discern whether they explained a significant part of the variance. Since there was only one deaccented referent in the entire data set, no statistical analysis was performed on this. Similarly, it was not possible to compute any statistics for H+L*, because there were too few items. This accent type occurred only once in the five-year-old group and not at all in the seven-year-old group, while it was used six times by the adults (four times for unused referents, and two times for brand-new referents). Of the seven other accent types, only L*+H showed a significant effect of AGE. The five-year-olds used this accent significantly less often than the adult group ($\beta = -1.43$, $z = -2.42$, $p < .05$), whereas the seven-year-olds did not differ from the adult group in this respect ($\beta = -0.62$, $z = -1.20$, $p = .22$). There also was no difference between the five- and the seven-year-olds ($\beta = 0.80$, $z = 1.29$, $p = .19$). However, there was no effect of TYPE for any of the accents, and there were no significant interactions.

These analyses have shown that in this data set, intonational marking of a new referent was not influenced by whether the referent is known (identifiable) for the speaker or not. If identifiability was to play a role, there should have been clear differences between the brand-new referents from chapter 4 (“a beaver”, “a seagull” etc.) and the unused referents in the present study (“Santa Claus”, “the sandman”). This was not the case. More importantly, there

5.4 Results and discussion

was not a single case of a deaccented identifiable referent, which would have been expected if identifiability due to cultural common ground had a similar effect on intonation as (immediate) previous mention (cf. chapter 4). New identifiable referents were marked in the same way as new unidentifiable referents. This result supports Lambrecht's (1994) hypothesis that the marking of identifiability is independent of the marking of activation. The fact that there was no interaction between AGE and TYPE furthermore indicates that already five-year-old children are aware of the difference between the two types of givenness, givenness due to general or cultural knowledge (identifiability), as in the case of the well-known characters, and marked by variations in the determiner, and givenness due to previous mention (activation), as marked by differences in intonation.

Part III: Methodological studies

Introduction

Up to this point, we have looked at the intonation of children and adults for topic-hood and information status, and at age-related differences in the phonetic realization of accents. However, as mentioned in the background, there are a number of methodological issues involved in research of this type. This last section of the dissertation addresses some of these issues.

First, claims on the mapping between intonation and information-structural categories are often based on results obtained from read speech. We have seen that in particular the results of adult speakers from chapter 4 are at variance with claims in the literature. It was argued that this discrepancy is in part due to the fact that the present data is drawn from unscripted speech, whereas previous experiments were based on read speech. However, it is also possible that the materials used to elicit the corpus are not directly comparable to those used by Baumann (2006) and Baumann and Grice (2006). In order to test the possible influence of speaking mode on intonation (i.e., spontaneous vs. read), the first chapter of this part reports the results of an experiment in which speakers read aloud text versions of the picture stories from chapter 4.

Another issue concerns the determination of phrase-boundaries in spontaneous child and adult speech. As mentioned in chapter 4, discerning where phrase boundaries are located in a given stretch of speech is not always easy. One possible way to approach the problem is to collect judgments from naïve listeners, who have been found to show considerable agreement in their perception of prosody in this respect. However, this method has so far not been used with spontaneous child speech. Chapter 7 therefore presents the results of a perception experiment investigating how untrained listeners perceive prosodic boundaries in the corpus of narrations, and evaluates the usability of this approach for child prosody analysis.

Third, it was argued repeatedly in this dissertation that it is essential to present acoustic evidence for the prosodic labels one is assigning, in particular when dealing with child speech, as it is otherwise unclear what those labels stand for. The preceding chapters have described one way of providing such

evidence. The final chapter of part III (chapter 8) now presents an alternative method: The nuclear contours are modeled mathematically, and whether accents that were labeled differently also differ from each other in their mathematical description is tested. This novel approach is applied to spontaneous child and adult language for the first time.

6 Information status marking in read speech

6.1 *Introduction*

The findings regarding information status marking obtained in chapter 4 differ quite strikingly from what one would have expected on the basis of the literature (Baumann & Grice, 2006; Baumann & Hadelich, 2003). Baumann and Grice (2006) argue that new referents should be marked with H* accents, while accessible referents should be marked with H+L* accents, and given referents by deaccentuation. However, different results emerged from chapter 4. In particular, for the adult group, there was a high number of L*+H accents in the new and the accessible condition, and overall only very few instances of H+L*, the accent that has been suggested to mark accessibility. Furthermore, although given referents had a tendency to be deaccented, still nearly 70% of them were accented. In chapter 4 it was argued that this divergence might in part be attributable to differences between read speech and natural speech. To test this hypothesis, I conducted an additional experiment with adult speakers, in which they read stories that verbalized what was shown in the pictures. If there are substantial differences between information status marking in spontaneous and read speech, this has important implications for the study of language development. Although children receive input also from read speech (e.g., fairytales), most of the time, they are exposed to spontaneous speech; therefore intonation patterns that occur in spontaneous speech will greatly influence the target model they are acquiring. This target model may in reality be very different from what we have assumed on the basis of previous studies that used scripted speech.

6.2 *Materials and procedure*

The picture stories were the same as described in section 4.2.2. In contrast to the previous experiment, the pictures were this time accompanied by text. Under each picture, a sentence was provided that described the picture. An example is provided in Figure 20. The target referent was always the last referring expression in the sentence, and in 10 out of 12 cases also the last word.

6.3 Participants

In two sentences, the target referent was followed by a verb particle (*aus* ‘out’ and *mit* ‘with’). Thus the sentences were of the same structure as the sentences that speakers had produced spontaneously. The full texts can be found in Appendix F. I assigned participants randomly to four different lists, each of which had a different trial order.

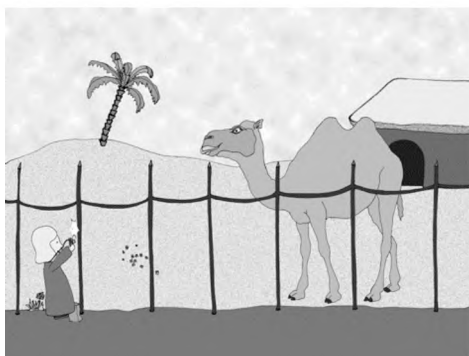


Figure 20: Picture from the *kamel* given story. The accompanying text was: “Das Mädchen nimmt seine Kamera und fotografiert das Kamel” ‘the girl takes her camera and takes a picture of the camel’.

6.3 *Participants*

Eight adult native speakers of German participated in the experiment (all female, mean age 22.5 years). They were all students of Linguistics or Psychology at the University of Potsdam and had no reported history of any speech or hearing deficits. None of the participants had taken part in the previous experiment. The participation was rewarded with a financial reimbursement or course credit.

6.4 *Data annotation, data selection and accent label consistency check*

Of the 96 items (8 participants x 4 referents x 3 conditions), 11 had to be excluded because of disfluencies or creaky voice. The remaining 85 items were annotated as described in section 4.3 and Appendix B4. The various accentua-

6.5 Results and discussion

tion types were again compared along a number of acoustic parameters. The results showed that their mean values differed significantly from each other in the way predicted from the GToBI literature (details of the analysis are provided in Appendix G).

6.5 Results and discussion

Almost all utterances (83) were produced with a low boundary tone. This is not surprising, as the punctuation in the texts (full stop at the end of each sentence) canonically suggests falling intonation. Table 14 shows the distribution of the different pitch accent types across the three experimental conditions.

		Status					
		New		accessible		given	
		N	%	N	%	N	%
Accent	deacc.	0	.0	5	17.9	26	92.9
	H*	7	24.1	2	7.1	0	.0
	L+H*	3	10.3	1	3.6	0	.0
	L*+H	0	.0	1	3.6	0	.0
	H+L*	9	31.0	9	32.1	0	.0
	!H*	0	.0	0	.0	0	.0
	H+!H*	10	34.5	8	28.6	2	7.1
	L*	0	.0	2	7.1	0	.0
	Total	29	100.0	28	100.0	28	100.0

Table 14: Distribution (absolute and relative) of pitch accent types by condition in the reading experiment.

The results are very much in line with the findings from literature. Previous perception experiments (Baumann & Hadelich, 2003; Chen et al., 2007) have

6.5 Results and discussion

found a strong preference for given referents to be deaccented. This strong preference is also borne out by the present data: The overwhelming majority (93%) of given referents was deaccented. Also, Baumann and Hadelich (2003) and Baumann and Grice (2006) had found that listeners perceive H+L* as an appropriate accent for accessible referents, and indeed, speakers realized referents in the accessible condition most often with this accent type. An unpredicted outcome is that H* accents were used less in the new condition compared to H+!H* and H+L* accents. In the perception experiments just mentioned, listeners had rated H* accents more appropriate than H+L* accents for new referents. However, in Baumann's (2006) corpus study, in which he analyzed the intonation of sentence-final referents in read-out newspaper texts, 69% of all new referents were also produced with either H+!H* or H+L* accents. This percentage is remarkably similar to the 66% found in the present data.

These results differ clearly from those found in the corpus of spontaneous narrations (cf. chapter 4), both in the distribution of boundary tones and in the distribution of pitch accent types. While in the spontaneous narrations the majority of utterances ended in a high boundary tone, in reading most utterances were low ending. Differing distributions of boundary tones across speaking modes have been reported before: For (Bari) Italian, Grice, Savino and Refice (1997) found that in spontaneously produced task-oriented dialogues, speakers produced questions mostly with L-% boundary tones, and only very rarely with a final rise (L-H%). In contrast, when speakers read transcripts of their dialogues aloud, they realized the majority of questions with a final rise. This and the present result show that written text, and in particular punctuation, superimposes strong restrictions on speakers' choice of boundary tones, which differs from what they typically use in unscripted speech.

Differences are also found in the distribution of pitch accent types. Table 15 compares the distribution in the read data with the distribution in the spontaneous data.

6.5 Results and discussion

		Status					
		new		accessible		given	
		mode		Mode		mode	
		read	spontan.	read	spontan.	read	spontan.
		%	%	%	%	%	%
Acc.	deacc.	.0	.0	17.9	8.7	92.9	32.6
	H*	24.1	21.6	7.1	13.0	.0	23.9
	L+H*	10.3	17.6	3.6	39.1	.0	4.3
	L*+H	.0	29.4	3.6	26.1	.0	8.7
	H+L*	31.0	3.9	32.1	4.3	.0	10.9
	!H*	.0	3.9	.0	.0	.0	2.2
	H+!H*	34.5	13.7	28.6	.0	7.1	13.0
	L*	.0	9.8	7.1	8.7	.0	4.3
	Total	100.0	100.0	100.0	100.0	100.0	100.0

Table 15: Distribution of pitch accent types by condition and mode. Differences of more than 20% between read and spontaneous speech are highlighted by bold face. N = 205.

The most striking difference between the two is the percentage of deaccented referents in the given condition. Whilst it is only about 33% in the spontaneous data, it is more than 90% in the read data. Secondly, while L*+H accents figured prominently in the adults' spontaneous production, they were virtually absent in read speech. Thirdly, we see that the accents H+!H* and H+L* were produced much more often in read speech.

The comparison of the two sets of results clearly shows that different intonation patterns are obtained depending on the task. Although the partici-

6.5 Results and discussion

pants saw the same picture stories with the target referents having the same information status as when the stories were told freely, we find that they use several accentuation types (including deaccentuation) in very different distributions. This shows that the differences between the results from the story-telling task reported in chapter 4 and the findings from the literature cannot be attributed to differences in the materials used. Rather, the results suggest that the intonation people use in read speech generally differs from intonation in spontaneous speech.

In light of the findings from chapter 4 (which showed that speakers tend to combine low pitch accent types with high boundary tones and vice versa), the low boundary tones induced by the punctuation may have discouraged the use of L^* and L^*+H accents. However, the distributions across speaking modes are different also for the other pitch accent types. In particular, the accents $H+L^*$ and $H+!H^*$ seem to be pervasive in read speech. Interestingly, Braun (2006) also reports that speakers produced mainly (!) $H+L^*$ nuclear accents in her reading experiment. Her study was not concerned with information status marking, but the elicitation of (contrastive and non-contrastive) topics and foci (cf. chapter 3). However, the participants read short paragraphs, a format comparable to the short stories in the present experiment. The finding that in these contexts most of the sentence-final words were typically realized with $H+L^*$ points to a special role of this accent in read speech. In chapter 4 it was speculated that $H+L^*$ accents might be part of a register used for narration. Both the present findings and the reports in the literature (Baumann, 2006; Braun, 2006) lend support to the idea that $H+L^*$ accents have to some extent become conventionalized elements of read-out narration and reporting. As such, adult speakers who are familiar with this style may also use it occasionally in (free) story telling, as observed in the present corpus.

While this explanation can account for the increased use of $H+L^*$ accents in the reading data, it fails to explain why there was also a much higher occurrence of deaccentuation. Why do reading intonation and spontaneous speech differ from each other in this respect? If deaccentuation can inform the listener about the given status of a referent, we should expect speakers to use this marking consistently, irrespective of mode. A consistent marking would

surely be beneficial for the communication process. Yet the results showed that deaccentuation was used considerably less often in spontaneous speech than in read speech. One possible explanation has already been mentioned in chapter 4: In spontaneous speech, functions other than information status marking may be more important, such as floor-keeping in conversation, flagging uncertainty or signaling attitudes towards the contents of the message or towards the interlocutor. Depending on the situation, the different functions of intonation will be more or less important. A stand-up comedian on stage is probably using intonation primarily to express how he feels towards what he is talking about rather than trying to keep the floor. In reading, many paralinguistic functions of intonation are likely to play only a minor role. Also, readers do not have to allocate as many resources to speech planning as a speaker in spontaneous discourse (see also Brown et al., 1980, p. 47). With “burdens” like these taken away from the speakers, they can concentrate on using intonation fully to reflect the information structure of the text, by deaccenting given referents, or by marking contrasting topics, for instance. Intonational information structure marking may therefore be especially pronounced in reading. What is more, written texts are often syntactically more complex and have a higher informational density than unscripted spoken language (Biber, 1988; Chafe, 1982). When reading such texts aloud, marking relations between parts of the sentence (e.g., topic-comment) or disambiguating syntax intonationally makes the text easier to process for the listener. As a consequence, using intonation in this way has probably become a codified element of “good reading”. One could perhaps even go so far to say that the kind of “typical” information-structure marking intonation scholars have postulated is to a certain degree a cultural skill that is learned in the course of literacy training. In fact, researchers working on the development of oral reading fluency have even remarked that, as children become skilled readers, their reading prosody takes on “a cultural normative character” (Schwanenflugel, Hamilton, Kuhn, Wisenbaker, & Stahl, 2004, p. 127). It is unlikely that reading intonation would be completely different from what people do in natural speech; but some aspects may have become exaggerated and conventionalized.

6.5 Results and discussion

The questions of why reading intonation may have developed in a certain way, and how exactly it deviates from spontaneous intonation are not the primary concerns of this study (but see Monschau, 2004 for a review of research on reading intonation in English). What the present experiment has demonstrated, however, is that there *are* clear differences in intonation between read and spontaneous speech. Since children learn language mostly from spontaneous speech, this means that the information-structural marking that is obtained in reading studies cannot straightforwardly be assumed to be the target model children are acquiring. Rather, it is necessary to also study more natural speech to determine what children are exposed to most of the time. We can then compare the realization of information-structural categories across different discourse types and look for those aspects that are invariant and children are likely to pick up on, such as the obligatory accentuation of new referents.

7 Naïve listeners' perception of prosodic boundaries

7.1 Introduction

Unlike the studies in the previous chapters, which were primarily concerned with the types of pitch accents adults and children use in different contexts, the experiment presented in this chapter is concerned with prosodic phrasing. In their investigations on information status marking of referents, Baumann (2006) and Baumann and Grice (2006) looked at phrase-final referents. To increase the comparability of their results and the present investigation, in chapter 4 I also restricted the analyses to phrase-final referents. For this it was necessary to identify the phrasal structure of the utterances. However, as has been mentioned, it was not always entirely clear where one phrase (IP) ended, and the next one began. This is because the phrasing of an utterance often correlates with syntactic structure, but prosodic structure is not isomorphic to syntactic structure; speakers insert breaks at points that do not correspond to major syntactic breaks, and may not reflect syntactic constituent structure in their prosody (Shattuck-Hufnagel & Turk, 1996; Steedman, 2000). For this reason, it is not possible to derive the prosodic structure of spoken utterances from their syntactic representation. Rather, prosodic phrasing must be determined on the basis of acoustic/phonetic cues. One way is to define a set of (measurable) criteria that must be fulfilled for a potential prosodic phrase boundary to qualify as such. Longer pauses (filled or unfilled) are probably the most obvious candidate for this, but there are other parameters such as increased segmental duration (pre-boundary lengthening) (Delattre, 1966; Klatt, 1975; Martin, 1970; Oller, 1973; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992) or pitch reset (e.g., Truckenbrodt, 2007). However, as will be explained below, we do not yet know how children mark prosodic phrase boundaries. It would therefore be difficult to decide which criteria to use. Alternatively, we can infer prosodic structure from the judgments of listeners – who are likely to use the available acoustic cues when they are processing speech (see Sanderman, 1996). In order to determine the location of

7.1 Introduction

phrase-boundaries in the data in chapter 4, the present study will take the latter approach.

Several studies have found that when naïve listeners are asked to indicate the presence of prominence on a word or the location of prosodic boundaries, their agreement is above chance (Buhmann et al., 2002; Mo, Cole, & Lee, 2008; Pijper & Sanderman, 1994; Streefkerk, Pols, & Bosch, 1997; Streefkerk, Pols, & Bosch, 1998). The agreement is typically reported using kappa coefficients, which are a statistical measure of the reliability of the agreement between raters. Values between 0.41 and 0.60 have been interpreted as “moderate agreement”, and values between 0.61 and 0.80 as “substantial agreement” (Landis & Koch, 1977). In the studies just cited, these kappa coefficients usually range between 0.45 and 0.72. However, the studies vary widely in the methodologies used, and some of the differences in agreement may be attributable to this. In some experiments, listeners transcribed read-out text (Pijper & Sanderman, 1994; Streefkerk et al., 1997; Streefkerk et al., 1998), which – as we have also seen in the preceding chapter – is in many aspects different from conversational speech. It does not require speech planning in the same way as spontaneous speech does, and punctuation (i.e., commas, brackets, full stops) imposes relatively strict constraints on the phrasal structure. Speakers typically produce fluent, clearly articulated and also clearly phrased sentences (Swerts & Heldner, 1996), in which listeners are likely to identify phrasal breaks without much difficulty and probably with high agreement. In other studies (Buhmann et al., 2002), transcribers had received training which arguably reduces the “naïvety”, were given additional information in form of a visual display of the sound files (waveform of the speech signal synchronized with the orthographic transcript), and did not listen to the sound files in real time, but had the possibility to listen to the files as often as they wanted.

More recently, researchers have begun setting up experiments that better reflect the situation in which language is naturally perceived, by using conversational speech material and asking listeners to give their judgments in real time and without the help of visual information (Mo et al., 2008). Since reasonable levels of agreement were achieved also in this setting, it has been argued that “naïve transcription is a valid method for prosody analysis which

can augment analysis solely based on expert labeling” (Mo et al., 2008, p. 735). However, the applicability of this technique for the analysis of child speech has not been explored.

This would be particularly interesting given that there is conflicting evidence regarding children’s ability to signal prosodic structure. Katz, Beach, Jenouri and Verma (1996) asked groups of five- and seven-year-old children and adults to describe different groupings of pink, green and white blocks, such as ‘(pink and green) and white’ vs. ‘pink and (green and white)’. While adults used word lengthening and pause duration to indicate the groupings, even the seven-year-old children failed to use pitch or duration to reliably indicate phrase structure. Katz et al.’s (1996) findings seem in conflict with results from a study by Dankovicová, Pigott, Wells and Peppé (2004), who found that adult listeners were able to pick up on children’s prosodic cues to disambiguate otherwise identical strings. This study is to my knowledge up to now the only investigation of adults’ perception of prosodic boundaries in (English) children’s speech. Minimal pairs like ‘chocolate-biscuits and honey’ vs. ‘chocolate, biscuits and honey’ were elicited from eight-year-olds by means of a picture book. The children’s productions were played to a group of adult listeners who indicated how many items they thought the speaker had described. The authors report that there was a high level of agreement among the listeners⁶², suggesting that the children produced reliable cues. Several explanations could account for the discrepancy between the Katz et al. (1996) and the Dankovicová et al. (2004) study. First of all, the children in Dankovicová et al.’s study were older than the ones in Katz et al.’s study (eight years vs. five years); it is plausible to assume that children’s ability to mark phrasal

⁶² Unlike most other studies, Dankovicová et al. do not report the kappa coefficient but a correlation measure (Kendall’s *W*, Kendall & Babington Smith, 1939), which is based on the raters’ judgement of their certainty that the speaker was referring to either two or three items (on a scale in six steps from *definitely two items* to *definitely three items*). Like Cohen’s Kappa, Kendall’s *W* ranges between 0 and 1, with 0 indicating no agreement, and 1 indicating perfect agreement. The value .74 thus suggests “substantial agreement” according to the scale suggested by Landis and Koch (1977).

7.1 Introduction

structure improves in the course of these three years. Second, Katz et al. (1996) did not test how listeners would interpret the children's utterances, so it is possible that the children might have used other, for listeners not perceptible cues, which Katz et al. did not measure acoustically. In this connection it is useful to recall Patel and Grigos' (2006) investigation of children's prosodic marking of question-/statement contrasts discussed in chapter 1. Patel and Grigos had found that four-year-old children manipulated other parameters in signaling questions than the older children: While the older children used F₀, the younger children relied on duration. It might be the case that the five-year-olds in Katz et al.'s (1996) experiment used cues other than the ones measured (e.g., intensity). Third, it is also possible that children find some contrasts easier to produce than others. In the Dankovicová et al. (2004) experiment, the children had to signal a lexical distinction (compound noun vs. sequence of two nouns), whereas the ambiguity in the Katz et al. (1996) stimuli was a syntactic one, marking different coordinate structures (Wells et al. 2004, p. 772).

Thus, we still do not know to what extent children younger than eight years of age produce prosodic cues to phrasal structure that are perceptible to (untrained) adult listeners. What is more, the studies just mentioned have looked at rather special cases in which phrasing was necessary to disambiguate. We know from Snedeker and Trueswell (2003) that adult speakers produce stronger cues if they want to disambiguate otherwise identical structures. It is possible that children produce clear cues if the situation requires disambiguation (as in the Dankovicová et al. study), but this does not necessarily mean that they produce similarly reliable cues when producing sentences with more complex syntactic structures in natural discourse. Finally, there has so far not been any investigation of naïve listener perception in German.

The present study sets out to investigate how untrained listeners perceive prosodic phrase boundaries in unscripted speech produced by German five- and seven-year-old children and adults, unaided by visual material, and in real time. The aim of the experiment is two-fold. First, it seeks to find out whether reasonable agreement on the location of boundaries can be obtained under these conditions. If this were the case, this would also provide evidence that

young children produce reliable cues. Second, if naïve listeners shared the perception of prosodic boundaries, their judgments could indeed be used to augment expert analysis, also with child data. Specifically, in the context of the study presented in chapter 4, the aim is to use the listener judgments to help decide in critical cases if a target word is phrase-final or not.

7.2 *Method*⁶³

Listeners were informed that speakers tend to organize their utterances into various “groupings” or “chunks” of words, especially when producing longer utterances. Their task was to indicate these groupings on a printed transcript of the stretches of sound they were going to hear. It was pointed out that the chunks do not necessarily correspond to the units that are delimited by punctuation in written language. The participants were instructed to put a vertical bar between words that they perceived as belonging to different “chunks” of the utterance. The participants listened to the recordings in real time over loudspeakers. Each sound file was played only once.

7.3 *Materials*

The materials for this perception experiment were taken from the corpus described in chapter 2. As explained in more detail chapter 4, the materials were designed in such a way as to elicit target referents in sentence-final (and therefore mostly also phrase-final) position, to allow comparison with Baumann’s (2006) and Baumann and Grice’s (2006) work. However, in about half of the utterances, speakers produced other material following the target word, and it was not always straightforward to decide if this material was still part of the phrase containing the target word, or not. In the sentence in (32), for instance, which was uttered by a seven-year-old, there was no clear pause after *Biber*, but it could not be excluded that some phrase-final lengthening may be present.

⁶³ I would like to thank Yoonsook Mo and Jennifer Cole for sharing information about their procedure with me.

7.3 Materials

- (32) Da dankt er dem Biber da der das Schiff zurück gebracht hat.
there thanks he the beaver since he the ship back brought has

'There he is thanking the beaver, because it has brought back the ship.'

For each age group (i.e., five- and seven-year-olds, and adults), sound files were excised from those utterances in which the target referent was not followed by a clearly perceptible pause ("critical utterances"). In addition, a number of other utterances, produced without disfluencies and typically consisting of only one or two clearly perceptible phrases (i.e., separated by a boundary tone and a clear pause), and which were therefore easier to label, were included in the experiment ("non-critical utterances") in order to present the listeners with a more balanced sample of what the speakers had produced.

Three different sets of materials were constructed, one from each age group. From the data of the five-year-old group, a total of 50 sound files were created from the narrations, varying in length from 4 to 15 words. The mean length was 7.8 words per excerpt. The excerpts came from 27 different subjects, resulting in an average of 1.8 items from each subject. From the seven-year-olds' narrations, 66 sound files were created (4 to 15 words, mean length 7.6 words), from 14 different subjects (4.7 items from each subject on average).

From the adult data, 61 sound files were created (5 to 19 words, mean length of 10.9 words), from 15 different subjects (4.06 items per subject on average). All sound files were normalized in amplitude using a root mean square (RMS) measure. Each sound file was followed by a pause of 5 seconds for listeners to write down their transcription. On the printed transcripts, words were capitalized according to German orthography, and separated by space, with no punctuation. Disfluencies and speech errors were contained in the transcripts.

7.4 *Participants*

Sixty native speakers of German (students from the University of Cologne, the University of Potsdam and the Radboud University Nijmegen, 42 female, 18 male, mean age 21 years) participated in the experiment, 20 listeners for each age group. None of the participants reported a history of hearing impairment. None had taken part in any of the other experiments.

7.5 *Analyses and results*

Two types of analyses were performed on the data, following procedures by Mo et al. (2008) and Cole et al. (2008). The first analysis evaluated how reliable the boundary marking was across transcribers. To this end, each interval between two words (hereafter transition) was marked with either “0” if the transcriber did not mark a boundary, or with “1”, if the transcriber put a vertical line to indicate a boundary. The second analysis determined a probabilistic boundary score for each transition between a target word and the following word. These scores were then used as an indicator to determine the location of IP boundaries in the critical cases from chapter 4.

7.5.1 Reliability analysis

A total of 27,740 transcribed transitions were subjected to the analysis. Multi-transcriber agreement studies like the ones cited in the introductory section typically use Fleiss’ Kappa (Fleiss, 1971) as a measure for agreement. However, there is a problem with this measure. Like other agreement measures such as Scott’s pi (Scott, 1955) and Cohen’s kappa (Cohen, 1960), Fleiss’ kappa assumes fixed marginals for agreement. That means that it is being assumed that raters know a priori how many cases they should assign to each category. But this is not the case in a study like the present one (and in fact also not in the studies cited above). When this constraint is being ignored, the resulting value of kappa can vary greatly depending on the distribution of cases in each category (i.e., symmetry), even if number of raters, categories, cases and percent of overall agreement are held constant (Randolph, 2005, p. 4). If there is great asymmetry in the distribution of cases across categories, this can lead to

7.5 Analyses and results

the so-called prevalence paradox: a low kappa value despite high agreement among the raters (see e.g., Brennan & Prediger, 1981). In the case of the present study, it seems fair to assume that listeners do not expect a boundary after every single word. Thus they are more likely to assume that there is no boundary between two given words than that there is a boundary. Hence we would expect an asymmetrical distribution of cases across the two categories “boundary”/“no boundary”. Fleiss’ kappa is therefore not an appropriate measure. Randolph (2005) introduced an alternative to Fleiss’ Kappa, the free-marginal multirater kappa (multirater κ_{free}), which circumvents the problem of the prevalence paradox. It is the measure that was used in the present study.

For each age group, free-marginal multirater kappa coefficients were calculated for all transitions (i.e., transitions in critical and in non-critical utterances), as well as for the subset of critical utterances only, and for the critical transitions (i.e., the transition between a target word and the following word). If the raters are in complete agreement then multirater $\kappa_{\text{free}} = 1$. If there is no agreement among the raters (other than the agreement that is due to chance), then multirater $\kappa_{\text{free}} \leq 0$. Table 16 displays the scores for each age group.⁶⁴

The overall scores and the scores for the critical utterances show the highest levels of agreement, with kappa coefficients between 0.70 and 0.79. This is in the range of “substantial agreement” according to the interpretation scheme proposed by Landis and Koch (1977). The agreement for the critical transitions was clearly lower, with scores for the five-year-old group and the adult group indicating “moderate agreement”, and scores for the seven-year-old group indicating “fair agreement”.

⁶⁴ The assumption that the distribution across categories would be asymmetrical was confirmed. In all three groups, the raters judged considerably more often that a transition did not constitute a boundary. Of the 6,820 transitions rated by the listeners in the five-years-group, only 1,222 transitions were marked as boundaries, compared to 5,598 which were judged as not constituting a boundary. Similar proportions are found in the seven-years-group (1,255 vs. 7,365) and in the adults-group (1,796 vs. 10,504). Thus the use of the free-marginal multirater kappa was justified.

	5-year-olds	7-year-olds	adults
Overall score	0.703	0.792	0.771
Critical utterances	0.715	0.734	0.768
Critical transitions	0.444	0.284	0.415

Table 16: Free-marginal multirater kappa scores for each age group: overall score and broken down by critical utterances and critical transitions.

In order to test for a potential training or fatigue effect that might have occurred in the course of the experiment, a Pearson's product-moment correlation between mean agreement per critical utterance and trial number was carried out for each group of listeners. If listeners were improving in the perception of prosodic boundaries by having heard more examples, one would expect there to be lower agreement in the beginning of the experiment, and higher agreement towards the end. Mean agreement and trial number would be positively correlated. Alternatively, the listeners' concentration might have declined over time, making their judgments more variable. This would lead to a negative correlation between mean agreement and trial number. The results of the tests show that there was no significant correlation in any of the three listener groups (listener group five-year-old group: $r = -.08$, $p = .68$; listener group seven-year-old group: $r = -.15$, $p = .38$; listener group adults: $r = -.09$, $p = .54$). The mean agreement in each group of listeners remained stable across the experiment.

7.5.2 Boundary scores

The reliability analyses have shown that overall, the raters showed substantial agreement regarding the location of prosodic boundaries in the data, irrespective of the age group the materials came from. However, the analyses have also shown that agreement for the critical transitions was lower than for the overall scores and the critical utterances. This indicates that despite substantial agreement overall, there is considerable individual variation in the perception of these critical transitions. Cole et al. (2008), using data from the Mo et al. (2008) study, applied a criterion of considering every transition that has a

7.6 Discussion

boundary score above the mean (i.e., mean agreement score over all items) as representative of perceived prosodic phrases. However, as the mean agreement scores are different for each age group, the present study used a different criterion: An agreement of at least 80% of the listeners (16 out of 20) on the presence of a boundary was required for a transition to be considered a boundary. Applying this criterion, 19 out of the 99 critical transitions were defined as marking a phrase boundary.

7.6 Discussion

In this experiment, untrained listeners were presented with natural speech from children and adults and asked to indicate where they perceived prosodic boundaries. The overall agreement scores (mean multirater $\kappa_{\text{free}} = .754$) are comparable to those from previous studies, and even somewhat higher. Mo et al. (2008), for instance, obtained an average agreement of $\kappa = .588$. However, it is possible that previous studies have underestimated rater agreement, since they did not use marginal-free versions of kappa. If we allow for some upward correction of their kappa values, the values in Mo et al.'s (2008) study and the values obtained in the present experiment are very comparable. This is quite remarkable given the fact that the listeners in the present study heard each item only once. In all other previous studies in which listeners had to mark boundary locations, the transcribers were allowed to listen to items repeatedly at their own pace (Buhmann et al., 2002; Mo et al., 2008). At the same time, the raters in the present experiment agreed much less on the presence or absence of boundaries in the critical transitions, indicating that these items were more ambiguous in their prosodic structure. The agreement in critical transitions was better for the youngest age group than for the other two, which could mean that these items were easier to rate. Note, however, that the overall scores were similar across age groups, which shows that the higher level of agreement in the critical transitions cannot be due to that listener group being "better" than the other two listener groups. Interestingly, the fact that listeners in the five-years-group had to deal with more variability in the material (the materials were drawn from more different speakers than in the other two groups) did not seem to have a detrimental effect. Unexpectedly, the

critical transitions in the seven-years-group had a lower score than in the other two age groups.

The fact that the overall level of agreement was substantial and that it did not seem to be related to age group warrants two conclusions. First, the method employed here is equally suitable for the analysis of adult speech and of child speech, and more specifically, of spontaneous child speech. Second, the children do not seem to be any less capable of signaling prosodic structure than the adults. Otherwise we would have expected more disagreement by the raters for the child groups compared to the adult group. This finding contrasts with the results from Katz et al.'s (1996) study, who had found that five-year-olds did not use F0 or duration to indicate phrase structure. But, as pointed out in the introduction, we do not know whether the children did not grasp the syntactic ambiguity and therefore failed to signal the correct structure, or whether the listeners would nonetheless have been able to hear phrasal breaks using other cues. The present finding that children are able to convey prosodic structure in their speech in a way that is perceptible by adults is in line with the Dankovicová et al. (2004) study on eight-year-olds, and suggests that this is true also for children as young as five years.⁶⁵ Furthermore, this experiment has shown that this extends to speech from free discourse. Note,

⁶⁵An alternative explanation for the high agreement among listeners is that they based their judgments primarily on syntactic considerations. Cole et al. (2008), for instance, report from their analyses of prosody perception of spontaneous American English that 45% of perceived boundaries comprise a major syntactic constituent, suggesting that syntax plays a role in boundary perception. However, since there is a strong correlation between syntactic and prosodic structure (e.g., Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991), this does not necessarily mean that listeners disregard prosodic information in their judgments. Furthermore, findings that non-native listeners without access to lexical, syntactic or semantic information in the speech signal predict upcoming prosodic boundaries on par with native listeners suggest that prosody is the most important cue in boundary perception (Carlson, Hirschberg, & Swerts, 2005). Yet without further analyses, it remains an empirical question whether listeners' judgments in the present experiment were driven more by syntactic or by acoustic and prosodic cues.

7.6 Discussion

however, that the material used here does not necessarily constitute a fully representative sample of child and adult speech. A large part of the material was explicitly selected because it was difficult to label prosodically, and the other part was selected because it seemed easier to label. Additional experiments with truly random samples are needed to confirm the findings.

What the present study has not addressed is the issue of *which* cues adults use to identify phrase breaks in child speech, and whether they are different from the cues listeners use when they are rating adult speech. Dankovicová et al. also ran acoustic analyses on the children's productions and related them to the listener ratings. They found that both final syllable duration and, to an even larger extent, pause duration were significant predictors for listeners' ratings. However, the children were already eight years old, and the material was of a rather special and restricted nature (disambiguation of otherwise identical structures). Thus the question remains open as to how young children mark prosodic boundaries in spontaneous speech, and which of the parameters they use actually cue adult listeners into the perception of these boundaries. Mo's (2008) findings for (adult) American English are interesting, as she observed that the phonetic cues most strongly associated with listeners' perception of a prosodic boundary were pre-boundary lengthening but also diminished intensity. This suggests that it could be useful to include both temporal and intensity measurements in future analyses.

8 Polynomial modeling of spontaneous child and adult intonation⁶⁶

8.1 Introduction

In the studies presented in this dissertation, intonation has been annotated using the GToBI framework. Empirical support for the labels used was sought by testing whether accents that were labeled differently also differed from each other along specific phonetic parameters, such as the alignment of F0 minima and maxima. However, the F0 turning points were determined manually, making this procedure rather time consuming. In addition, it is not always easy to determine the location of pitch minima and maxima (e.g., because of segmental effects), and this uncertainty has consequences for the reliability of both the alignment and the pitch range measurements. When in doubt, labelers who have to decide on the location of a turning point may (without being aware of it) be inclined to select a location that is favorable to the prosodic label they have assigned, which potentially adds a certain amount of subjectivity. In view of these difficulties, it would be desirable to have an additional, more objective way of testing whether the labels somehow correspond with measurable properties of the speech signal.

Such empirical acoustic evidence for intonation labels in English has been presented by Grabe, Kochanski and Coleman (2007). They modeled the fundamental frequency (F0) of hand-labeled accents mathematically using polynomial equations, and showed that the accents that were assigned different labels were also significantly different from each other in their mathematical descriptions. Polynomial equations are a way to describe curves, or rather continuous functions – of which F0 is assumed to be one as well – in a mathematical expression constructed from variables and constants (e.g., $3x^2 + 4x + 5$). They are one of a number of approaches used in speech technology to fit F0

⁶⁶ A version of this chapter will appear as De Ruiter, L.E. (in press), Polynomial modeling of child and adult intonation in German spontaneous speech, *Language and Speech*.

8.1 Introduction

curves. Other curve-fitting models include the Fujisaki model (Fujisaki, 1992), MOMEL (Hirst, Di Cristo, & Espresser, 1993) or Tilt (Taylor, 2000).

The corpus used in the Grabe et al. (2007) study consisted of 714 read-out sentences, produced by 42 speakers. The nuclear contours of these sentences (i.e., the final pitch accent in the phrase and the subsequent boundary tone) were manually annotated according to the IViE (Intonational Variation in English) labeling system (Grabe, 2002), which is an autosegmental-metrical transcription system based on ToBI, but developed to allow dialect-independent transcription of English intonation (for details, see also Grabe, 2004). Seven different nuclear contour types were found in the corpus and modeled with orthogonal Legendre polynomials (details are given in section 8.2.2 below), resulting in a concise mathematical description of each accent. Statistical analyses showed that parameters of the polynomial descriptions of six of the seven contour types that had been identified by the labelers differed significantly from each other in at least one of the three coefficients. The authors conclude that polynomial modeling can provide intonational phonologists with a tool to empirically validate linguistic descriptions of intonation (Grabe et al., 2007, p. 299).

Against this background, the present study sets out to investigate whether polynomial modeling can be applied to the present corpus to gain additional empirical support for the prosodic labels assigned here. Note that unlike Grabe et al.'s corpus, the speech material analyzed here consists of natural rather than scripted speech. What is more, the speakers in this study were drawn from three different age groups: five-year-olds, seven-year-olds and adults, adding more variability to the data in terms of speaking rate and average pitch. These features make the data potentially more difficult to model.

In addition, this chapter presents an extension to the polynomial model that shows how information about the relative alignment of tonal targets with the segmental string can be derived from the modeled curves, connecting this new approach directly with the research on tonal alignment (e.g., Arvaniti, Ladd, & Mennen, 1998; Atterer & Ladd, 2004). Finally, I explore how polynomial modeling may provide a window into intonational development.

8.2 Method

8.2.1 Data and annotation

The data set consists of 291 prosodically annotated utterances taken from the corpus of narrations described in chapter 2 and in more detail in chapter 4. These utterances were the same as the ones used for the pitch accent label consistency check in chapter 4 (see Appendix B). This means that they were utterances in which the target referents (*Möwe* (/ˈmø:və/, ‘seagull’), *Biber* (/ˈbi:bə/, ‘beaver’, *Biene* (/ˈbi:nə/, ‘bee’, and *Kamel* (/kaˈmɛ:l/, ‘camel’) occurred phrase-finally. Two examples are provided in (33) and (34).

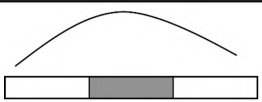
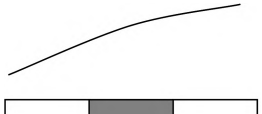
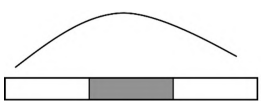
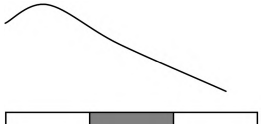
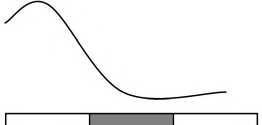
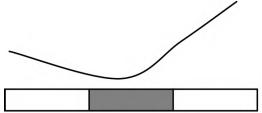
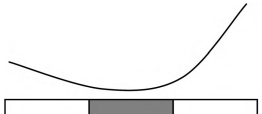
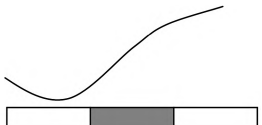
(33) Er malt eine Biene.
‘He draws a bee.’

(34) Sie kommt wieder zur Möwe.
‘She comes back to the seagull.’

As reported in chapter 4, seven different nuclear pitch accent types and five different boundary tones were labeled in the data. In addition, a substantial number of items were deaccented. Not all possible combinations of pitch accents and boundary tones occurred, and some combinations occurred only very rarely (see Appendix B). The boundary tones H-% (high boundary) and H-^H% (upstepped boundary) were subsumed under one (simplified) label, H%. The boundary tones L-% (low boundary) and !H-% (downstepped boundary) were subsumed under the label L%. Table 17 displays the pitch accent + boundary tone combinations that occurred at least ten times. (Following Grabe et al. 2007, accent labels such as H*, and boundary tone labels, such as L%, are separated by commas in the transcription.) For brevity, these combinations will be referred to as nuclear contours, whose shapes are shown in Table 17. The first column gives the GToBI labels, the second column provides a stylized representation of the typical F0 shape, and the third column

8.2 Method

gives the frequency of the accent in the data set. Note that the category of downstepped accents !H*,L% was not collapsed with its non-downstepped counterpart (which was done by Grabe et al., 2007). With eleven different nuclear contours, the data set contains a larger number of contours than was modeled in the Grabe et al. study, where only seven different nuclear contours were analyzed. This is due to the fact that I also included deaccented items in the data, and also to the fact that the GToBI system also contains right-headed accents (e.g., L+H*), which do not exist in the IViE transcription system, where an accent like L+H* would simply be described as H*.

Accent label	Stylized F0	N
!H*,L%		16
H*,H%		19
H*,L%		40
H+!H*,L%		34
H+L*,L%		10
L*+H,H%		47
L*,H%		27
L+H*,H%		23

8.2 Method

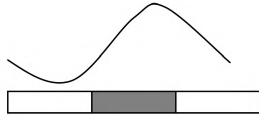
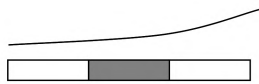
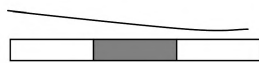
L+H*,L%		24
deacc.,H%		16
deacc.,L%		29
other		6
Total		291

Table 17: Autosegmental-metrical contour labels, F0 stylizations and frequency of occurrence (N) of the accents observed in the data set. The grey-shaded area indicates the lexically stressed syllable; the white areas preceding and following it represent the pre- and post-accentual syllable, respectively.

8.2.2 Polynomial modeling

In this section, first the polynomial approach by Grabe et al. (2007) is described. I will then give details of the way polynomial modeling has been implemented in this study, and point out where the method differs from the approach taken by Grabe et al.

In their study, Grabe et al. analyzed the region of F0 beginning 100 milliseconds before the nuclear accent of each sentence in the corpus, and extending to the end of the voiced part of the IP. All utterances in their corpus were designed to be fully voiced, but some voiced fricatives tended to be devoiced phrase-finally, so that some utterances contained unvoiced material. These parts were discarded from the analysis. In addition to measuring F0, Grabe et al. also extracted measures of loudness and aperiodicity from the signal (details can be found in Kochanski, Grabe, Coleman, & Rosner, 2005). In the course of fitting the polynomial model to the data, these measures were later

used to give more weight to loud and sonorant regions, such as syllable centers, assuming that in these regions, F0 measures are more reliable and F0 movements may be more perceptually relevant (Grabe et al., 2007, p. 287). Before fitting, all F0 values were normalized by dividing them by the speaker's mean, and subtracting 1. Thus, a normalized F0 of zero corresponded to the speaker's average F0. Furthermore, Grabe et al. shifted and scaled the time axis so that the nuclear contour region spanned values between -1 and 1, which is a prerequisite for Legendre polynomials. The F0 data in the analysis region of each sentence was then modeled as a best-fit sum of Legendre polynomials (the exact details of the procedure are given in Kochanski et al., 2005). Legendre polynomials belong to the class of orthogonal polynomials. As argued by the authors, orthogonal polynomials have the advantage of minimizing correlations among the coefficients, which would otherwise have to be taken into account in the statistical analysis. In contrast to other orthogonal polynomials, Legendre polynomials ensure that the coefficients are equally sensitive throughout the utterance (ibid.). After fitting, each nuclear contour is described by a model. This model is specified by a set of coefficients (c_0, c_1, c_2, c_3) that multiply the different Legendre polynomials before they are added together:

$$\text{Equation 1} \quad M(x) = c_0 + c_1 \cdot x + c_2 \cdot \left(\frac{1}{2}(3x^2 - 1)\right) + c_3 \cdot \left(\frac{1}{2}(5x^3 - 3x)\right)$$

The polynomial analysis bears similarities to a Fourier analysis in that the lower-ranking polynomials identify the more slowly varying properties of the curve, whereas the higher-ranking polynomials identify the more rapidly changing properties (Grabe et al., 2007, p. 289). The more complex a curve is, the more polynomials are needed to fully describe it. As noted by Grabe et al., the first four polynomials ($c_0 - c_3$, displayed in Figure 21) have straightforward physical interpretations:

8.2 Method

- The first coefficient, c_0 , corresponds to the average F0 of the accent after pitch normalization.
- The second coefficient, c_1 , gives an indication of the overall slope (falling or rising) of the accent.
- The third coefficient, c_2 , specifies the overall curvature of the accent, which can be a broad dip or a rise in the middle of the accent.
- The fourth coefficient, c_3 , corresponds to a wave-like shape.

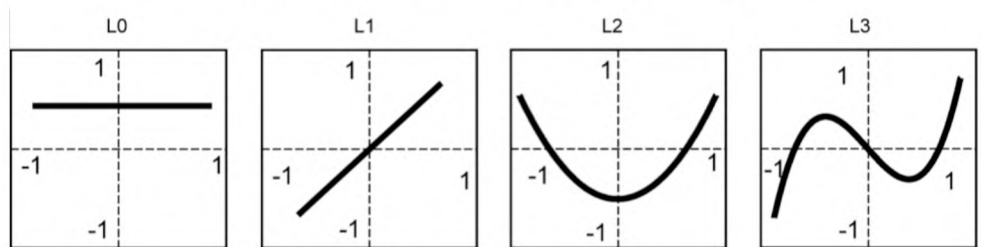


Figure 21: The first four Legendre polynomials L_0 - L_3 . Following the naming convention introduced by Grabe et al. (2007), the first coefficient (c_0), will be referred to as AVERAGE, the second coefficient (c_1) will be referred to as SLOPE, the third coefficient (c_2), as PARABOLA, and the fourth coefficient (c_3) as WAVE.

The following paragraphs now describe in detail how the nuclear contours in the present data set were modeled.

Prior to the analysis, the F0 tracks were inspected for errors such as octave jumps and manually corrected where necessary. The analyzed domain consisted of the (voiced regions of the) lexically stressed syllable, the pre-tonic syllable, and the post-tonic syllable (see below for details on how the voiced region was determined). This is different from Grabe et al.'s domain (see above). There were two reasons to define the region differently: First, Grabe et al.'s criterion is based on adult speech, while the present data set also comprises child speech. Young children typically have a lower speech rate, which increases with age until at least age eleven (e.g., Boutsen & Hood, 1996; Sturm & Seery, 2007). This means that for the young speakers, a starting point of 100 milliseconds preceding the centre of the stressed syllable would still be way into the stressed syllable. As a consequence, the analysis region would not

even cover the entirety of the stressed syllable, which is clearly undesirable. Second, unlike the IViE transcription system, the GToBI scheme also contains right-headed accents (e.g., L+H*), which means that the pitch movement on the pre-tonic syllable is considered to be an important part of the overall accent shape for some accent types. Any criterion of having the beginning of the analysis region start at some arbitrary distance from the center of the stressed syllable (e.g., 200 milliseconds) risks “cutting off” parts of the pre-tonic syllable even with adult speakers. The modeled F0 curve would consequently not be a truthful representation of the F0 curve that the intonational labeling was based on, and could therefore not be used to find empirical support for the accent labels. Determining the analysis region on the basis of the syllable structure avoids these problems. An illustration of the analysis region is provided in Figure 22. Within the analysis region, F0 was measured in steps of 5 milliseconds. At each point at which F0 was measured, intensity and harmonics-to-noise ratio⁶⁷ (Boersma, 1993) were also extracted. These two measures were later combined in a weighting parameter used in the fitting algorithm described below.

⁶⁷ Harmonics-to-noise ratio (HNR) quantifies the amount of additive noise in the speech signal. Additive noise can arise from airflow turbulences occurring at the vocal folds during phonation: When the vocal folds are not completely closed – as is the case in irregular phonation like creaky voice – air passes through the vocal folds and causes turbulence. This results in friction noise, which is in turn reflected in a higher noise level (Ferrand, 2002, p. 281). Hence, HNR can be used as an indicator of the periodicity of the speech signal.

8.2 Method

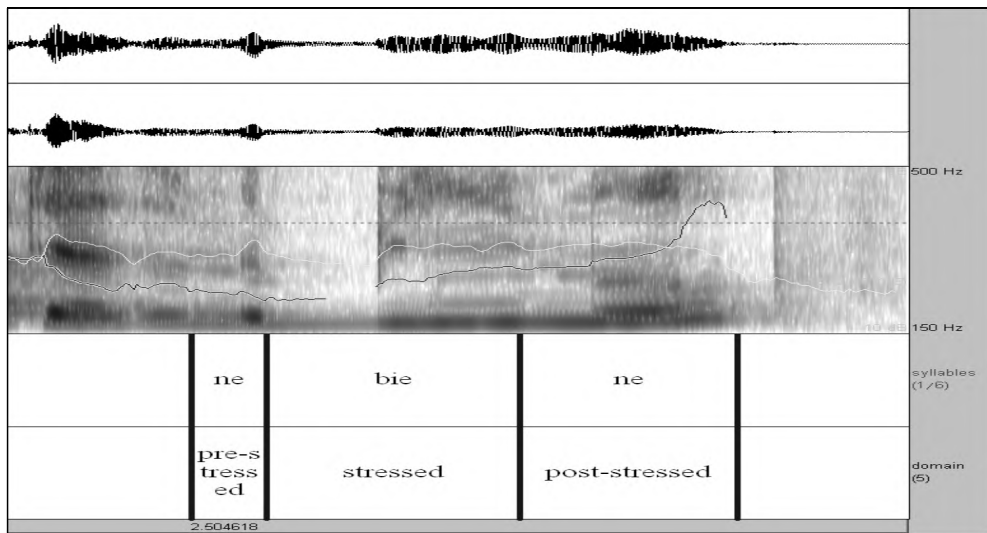


Figure 22: Amplitude envelope, spectrogram, intensity (thin light line) and F0 (thick black line) curve of an utterance by a five-year-old, *Guck, der malt ne Biene* ‘Look, he is drawing a bee’. The domain in which F0 is measured consists of the pre-tonic syllable (“ne”), the lexically stressed syllable (“bie”) and the post-tonic syllable (“ne”).

Before fitting the data, F0 and time were normalized. For F0-normalization, all F0 values were divided by the speaker’s mean (which was calculated by taking the average of all first unstressed syllables in all utterances made by the speaker), and subtracting 1 from it. Thus, a value of 0.1 corresponds to an F0 that is 10% above the speaker’s mean. For the time normalization, the time axis of the analysis domain was shifted and scaled to values between -1 and 1.

Like in Grabe et al., the model was specified by a set of four coefficients, c_i , that multiply the different Legendre polynomials before they are added together (cf. Equation 1). For the estimation of the coefficients of the Legendre polynomials that best describe a given intonation contour, I used Polyfit (De Ruiter, 2008), a customized computer program written in C++.

The program reads in normalized F0 values and a weighting parameter (described below) and calculates those Legendre coefficients that minimize

the difference between the predicted polynomial and the original pitch contour as estimated by Praat's pitch tracking algorithm.⁶⁸ The weighting parameter w is used to de-emphasize regions that are not important for the overall shape of the contour (such as microperturbations that are due to the segmental structure), and to give more weight to perceptually important loud and sonorant regions such as vowel centers (Kochanski et al., 2005, p. 1043). A higher w for a certain time window forces the algorithm to model F0 values in this region with more precision. The weighting parameter used in this study is a combination of the intensity and the HNR of the signal. Intensity was normalized by dividing each value by the mean intensity of the voiced parts of the entire utterance. Unlike intensity, HNR values usually cover a wider range of values and can also be negative, in cases where there is more noise than harmonics in the signal (e.g., in voiceless regions). I normalized the HNR measures using a sigmoid function, which transforms all possible values (from ∞ to $-\infty$) into values between 0 and 1. Hence negative HNR values (where there is a lot of noise in the signal) receive a low score near 0, whereas positive ones receive a score closer to 1. I calculated the parameter k in Equation 2 below using the criterion that a HNR value of 15dB (roughly equal to 97% energy from the harmonic part, see Praat manual on "harmonicity") receives an H-score of 0.75. The resulting coefficient k is 0.02453.

⁶⁸ The quantity that is minimized is a chi-square related merit function:

$$m = \frac{1}{N} \sum_{i=1}^N \left[\frac{\text{data}_i - \text{pred}_i}{w_i} \right]^2$$

where w_i is a weighting quantity indicating the relative contribution of data point i to the merit function. The program uses a General Linear Least Squares algorithm based on normal equations and Gauss-Jordan elimination, and is described in detail in Press, Teukolsky, Vetterling and Flannery (1988).

8.3 Results

$$\text{Equation 2} \quad H(t) = \frac{1}{1 + e^{-kt}}$$

The weighting parameter w was then the product of normalized intensity (I) and the standardized HNR value (H):

$$\text{Equation 3} \quad w = I \cdot H$$

As described above, the analysis region was a three-syllable-domain around the lexically stressed syllable. However, unvoiced regions (like devoiced vowels) at the beginning or the end of the domain can be problematic. When the program determines the coefficients to model the intonation contour, it mainly fits the polynomials to the voiced parts while the polynomials can take any form for unvoiced regions. This is not harmful for voiceless regions in the middle of a voiced region, if we assume that the F0 contour constitutes a smooth function; the algorithm interpolates between the voiced regions. However, for voiceless regions before pitch onset or after pitch offset, the fitting becomes unpredictable. Note that weighting alone cannot solve this problem, as a very low w would still “allow” the program to fit almost any curve. To avoid this problem, I set the domain to start at the first voiced frame within the original three-syllable domain, and to end at the last one. Following this approach, the analysis domain was adjusted for 211 out of the 291 nuclear contours. For these contours, the domain was on average shortened by 10% of the overall duration of the three-syllable domain.

8.3 Results

Figure 23 shows the original F0 curve and the modeled curve of two contours superimposed onto each other. One can clearly see that the microprosodic effects such as the pitch depression in the intervocalic voiced obstruent [b] of the word *Biene* in the left panel have not been modeled.

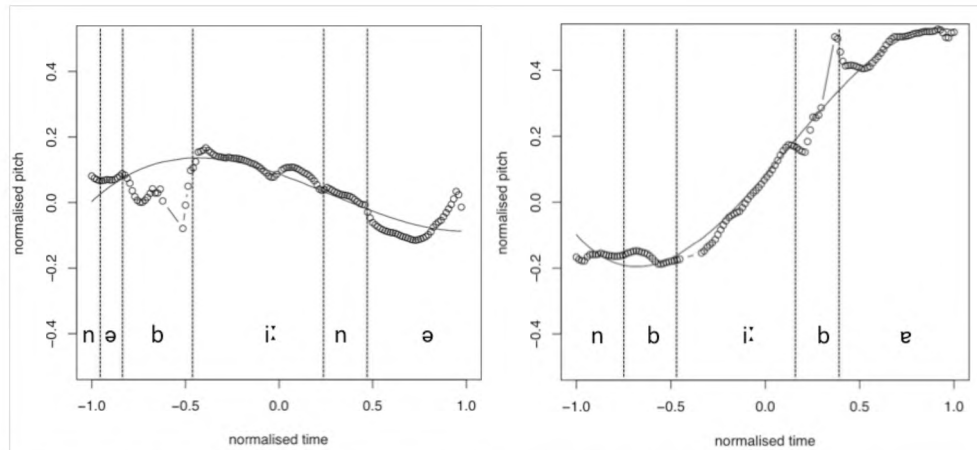


Figure 23: Examples of the (normalized) original F0-track (circles) and the curve modeled by Polyfit (continuous line). The x-axis plots the normalized time (from -1 to 1), the y-axis shows the units of the normalized F0. The panel on the left is an example of an accent that was labeled H*,L%, the panel on the right is an example of an accent that was labeled L*+H,H%.

Before presenting the results of the statistical analysis, I will first show two selected examples of contour profiles, following Grabe et al. (2007). These profiles can make the results of the contour pair contrasts easier to interpret. Figure 24 shows the contour profiles of the two most frequent nuclear contours, H*,L% and L*+H,H%, together with stylized shapes of these contours, which are based on the mean values for the four coefficients of each contour. The profiles of the other contours can be found in Appendix H. If c_0 is negative, this indicates that the mean F0 for this contour is low, while a positive c_0 indicates a mean F0 that is higher compared to the average F0 of the speaker. A negative c_1 is a sign of a predominantly falling accent, whereas a positive second coefficient means that the slope is mainly rising. If c_2 is negative, the F0 curve is convex (dome-shaped), if it is positive, the curve is concave (cup-shaped). Finally, c_3 describes whether the overall shape is falling-rising-falling (if it is negative), or rising-falling-rising (if it is positive).

8.3 Results

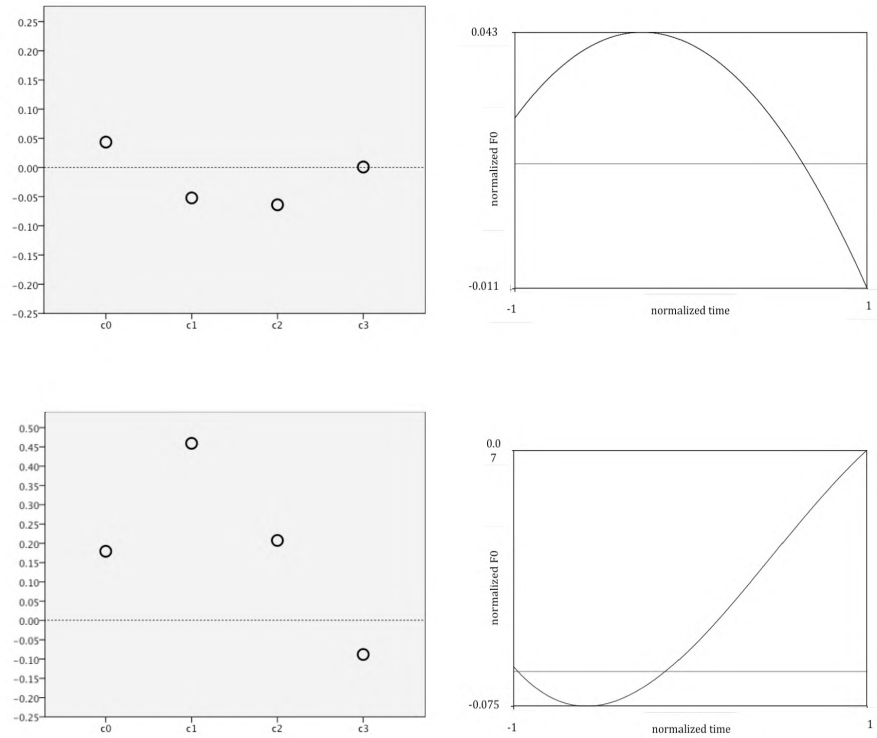


Figure 24: Four-coefficient F0 profiles and stylized contour shapes for H*,L% contours (top panels) and L*+H,H% contours (bottom panels). The four-coefficient F0 profiles are shown on the left hand side, with normalized values of the coefficients (c0-c3) plotted on the y-axis. The stylized contour shapes are shown on the right hand side, with normalized time plotted on the x-axis and normalized F0 plotted on the y-axis. The horizontal line indicates the speaker mean F0.

The statistical analyses were performed using R version 2.6.2. Instead of a Multivariate Analysis of Variance (MANOVA), which was used by Grabe et al. (2007), I tested for differences between the accents using linear-mixed effect (LME) models. Four LME models were constructed, one for each of the four coefficients. The models tested whether the contour label (henceforth: CONTOUR) is a significant predictor for a given coefficient. The models specified

the coefficient as dependent variable, CONTOUR as fixed factor. However, the items are not all of the same structure. The majority of items (N = 167) have an open lexically stressed syllable and penultimate stress (e.g., *Biber* (/ˈbiːbɐ/), but in the *Kamel* items (N = 118), the lexically stressed syllable is closed, and stress occurs on the final syllable of the word. Syllable structure and stress pattern of a word affect the shape of the pitch contour. For German, Möbius and Jilka (2007) found that in falling contours, the peak occurs earlier in closed syllables and when the nuclear accent occurs on the last syllable of an IP. These two conditions are both met by the *Kamel* items. We may therefore expect that the contours of the *Kamel* items differ in their shape from the other contours, and consequently also in their coefficients. For this reason, I included STRESS (with the two levels *penultimate* and *ultimate*) as an additional fixed factor in the models. In a first step, full models including both predictors (CONTOUR and STRESS) and their interaction were specified. Predictors with a p-value larger than 0.1 were removed if this did not deteriorate the fit of the models (as estimated by a likelihood-ratio test). Both SUBJECT and ITEM were initially included as crossed random factors. However, in three of the four models (for AVERAGE, SLOPE, and WAVE), the variance explained by ITEM was effectively zero, and this factor was therefore removed from the model. All reported p-values were obtained by Laplace approximation.

The results show that in all models, CONTOUR was a significant predictor for the four coefficients. With the exception of four nuclear contour pairs, all contours differed from each other in at least one coefficient. The four contour pairs for which no statistically significant differences were observed were: 1) $\emptyset, L\%$ and $!H^*, L\%$ 2) $\emptyset, H\%$ and $H^*, H\%$ 3) $L+H^*, L\%$ and $H^*, L\%$ 4) $H+!H^*, L\%$ - $!H^*, L\%$. There was no significant main effect of STRESS, but a significant interaction with CONTOUR in three of the four models (AVERAGE, SLOPE, and PARABOLA). For three nuclear contours ($L+H^*, H\%$ and $L^*+H, H\%$ and $H+!H^*, L\%$), items with ultimate stress showed a different pattern from items with penultimate stress. This did, however, not have an influence on whether these nu-

8.3 Results

clear contours differed significantly from other nuclear contours or not. All results are summarized in Tables 18 to 21.⁶⁹

⁶⁹ There was no difference between the three age groups with respect to the mean error (a measure of the distance between model and fit) of the curves modeled by Polyfit. The mean error for the adult group was 0.12, which was not significantly different from the mean error in the five-year-old group (0.15, $p = .65$) and the mean error in the seven-year-old group (0.08, $p = .45$). This was determined by means of a LME model with mean ERROR as dependent variable, AGE as fixed factor and WORD as random factor. The two younger age groups did not differ significantly from each other, either ($p = .24$)

8.3 Results

Tables 18 to 21: Half-matrix displaying contour pairs that are significantly different from each other in terms of AVERAGE, SLOPE, PARABOLA, and WAVE. Black cells indicate differences significant at $p \leq .01$, grey cells indicate differences significant at $p \leq .05$.⁷⁰

AVE-RAGE	∅,L %	∅,H %	L+H*,L %	L+H*,H %	L*,H %	L*+H, H %	H+L*,L %	H+!H*, L %	H*L, %	H*,H %	!H*L %
!H*,L%		Black	Grey	Black	Black	Black	Black	Black	Black	Black	Black
H*,H%	Black		Black	Black	Black	Black	Black	Black	Black	Grey	
H*,L%	Black	Black		Black	Black	Black	Black	Black	Black		
H+!H*, L%				Black	Black	Black	Black	Black			
H+L*,L %		Black		Black	Black	Black	Black				
L*+H,H %	Black		Black		Black						
L*,H%	Black	Grey	Grey	Black							
L+H*,H %	Black		Black								
L+H*,L %		Black									
∅,H%	Black										
∅,L%											

Table 18

⁷⁰ For reasons why Bonferroni correction and other p-value adjustments are inappropriate for this type of data, see Gelman, Hill and Yajima (2009), and Gelman and Tuerlinckx (2000).

8.3 Results

SLOPE	∅,L %	∅,H %	L+H*,L %	L+H*, H%	L*,H %	L*+H, H%	H+L*,L %	H+!H*, L%	H*L, %	H*,H %	!H*L %
!H*,L%											
H*,H%											
H*,L%											
H+!H*, L%											
H+L*,L %											
L*+H,H %											
L*,H%											
L+H*,H %											
L+H*,L %											
∅,H%											
∅,L%											

Table 19

PARA- BOLA	∅,L %	∅,H %	L+H*,L %	L+H*, H%	L*,H %	L*+H, H%	H+L*,L %	H+!H*, L%	H*L, %	H*,H %	!H*L %
!H*,L%											
H*,H%											
H*,L%											
H+!H*, L%											
H+L*,L %											
L*+H,H %											
L*,H%											
L+H*,H %											
L+H*,L %											
∅,H%											
∅,L%											

Table 20

8.3 Results

WAVE	∅,L %	∅,H %	L+H*,L %	L+H*, H%	L*,H %	L*+H, H%	H+L*,L %	H+!H*, L%	H*L, %	H*,H %	!H*L %
!H*,L%											
H*,H%											
H*,L%											
H+!H*, L%											
H+L*,L %											
L*+H,H %											
L*,H%											
L+H*,H %											
L+H*,L %											
∅,H%											
∅,L%											

Table 21

In what follows, I will describe the statistical results for the eleven different contours. In keeping with Grabe et al. (2007), I assume that one significant difference between two contours is sufficient evidence that the two accent patterns that were labeled differently do have different F0 manifestations. Note, however, that also in my data, in most cases the contour pair comparisons brought about more than one statistically significant difference (see Table 22).

8.3 Results

	un- acc,L %	un- acc,H %	L+H*,L %	L+H*,H %	L*,H %	L*+H,H %	H+L*,L %	H+!H*,L %	H*L, %	H*,H %
!H*,L%	0	3	3	4	3	4	1	0	1	1
H*,H%	2	0	3	2	3	2	3	2	3	
H*,L%	2	3	0	4	2	4	3	3		
H+!H*,L %	1	2	2	4	3	4	1			
H+L*,L%	1	2	3	3	3	4				
L*+H,H%	4	3	4	1	4					
L*,H%	3	4	4	4						
L+H*,H%	3	2	4							
L+H*,L%	1	2								
un- acc,H%	2									
unacc,L%										

Table 22: Half-matrix showing number of statistically significant differences in the coefficients between the nuclear contours pairs. Light-grey shading indicates that there was one significant difference between two nuclear contours, dark-grey shading indicates that there was more than one.

In order to keep the descriptions within reasonable length, not all 55 contour pair comparisons will be described. I will restrict the descriptions to a few comparisons for each contour, following the structure of the half matrices in Tables 18 to 21 (i.e., going in columns from left to right). (For ease of orientation, in each of the following paragraphs the accent under consideration will be set in boldface.)

First, I discuss the two contours that were not included in the Grabe et al. study, namely the two deaccented contours. Deaccented items that were followed by a low boundary (**∅,L%**) had an F0 which was on average lower than the speaker mean, which distinguished them for example from deaccented items that were followed by a high boundary (**∅,H%**), which had a higher mean F0. From most other contours, items labeled **∅,L%** were distinguished by their shallow SLOPE and slightly concave PARABOLA. However, the differ-

ences between $\emptyset, L\%$ and $!H^*, L\%$, and between $\emptyset, L\%$ and $!H^*, L\%$ were not significant.

Deaccented words with high boundary tones ($\emptyset, H\%$), in turn, had a slight rising SLOPE and were slightly cup-shaped (positive PARABOLA). With this, they differed from most other contours, whose SLOPE and PARABOLA were typically more pronounced. However, no significant differences were found between $\emptyset, H\%$ contours and contours labelled H^*, H .

Moving on to the first “genuine” nuclear contour, $L+H^*, L\%$, we see that this contour had a dome-shaped (negative PARABOLA). This was sufficient to distinguish it from clearly cup-shaped rising contours like $L^*+H, H\%$ and $L^*, H\%$, but also from other more dome-shaped contours such as $H+!H^*, L\%$, which had a negative PARABOLA as well, but was not as strongly dome-shaped as $L+H^*, L\%$. However, in none of the coefficients did $L+H^*, L\%$ turn out to be significantly different from $H^*, L\%$.

$L+H^*, H\%$ is a contour which had a steeply rising SLOPE. With this, this contour type differed significantly from $H^*, H\%$ contours whose SLOPE was less steep, and from $H+!H^*, L\%$ and $H+L^*, L\%$, whose SLOPES were mostly falling. From other contours $L+H^*, H\%$ contours differed mainly with respect to PARABOLA: $L+H^*, H\%$ contours were cup-shaped, and differed significantly from and $L+H^*, L\%$, which had a domed shape, and from $L^*, H\%$ and $L^*+H, H\%$, which were also cup-shaped, but more strongly curved. $L+H^*, H\%$ contours were one of the contours for which a significant interaction with STRESS was found. In items with ultimate stress, these contours had a somewhat lower AVERAGE and a steeper overall SLOPE.

With their narrowly cup-shaped PARABOLA, $L^*, H\%$ contours differed significantly from falling contours such as $H^*, L\%$, which were dome-shaped, and other rising contours like $H^*, H\%$ and $L^*+H, H\%$, which were cup-shaped as well, but not as narrow.

Like the other rising contours, $L^*+H, H\%$ contours exhibited a steep rising SLOPE, distinguishing them from falling contours with their predominantly falling slopes. $L^*+H, H\%$ contours were differentiated from the rising contour $H^*, H\%$ in terms of PARABOLA: $L^*+H, H\%$ contours had a more pronounced cup-shape. Furthermore, $L^*+H, H\%$ contours showed a significant interaction

8.3 Results

with STRESS in PARABOLA. In words with ultimate stress, the cup-shape was narrower. There was also a marginally significant interaction ($p = .05$) with STRESS in SLOPE. In words with ultimate stress, the SLOPE in $L^*+H,H\%$ contours was more steeply rising than in words with penultimate stress.

Turning to $H+L^*,L\%$ contours, we see that their steeply falling SLOPE differed significantly from $H^*,H\%$ contours, whose SLOPE was predominantly rising, and from $H^*,L\%$ contours, whose SLOPE was also falling, but less steep. Furthermore, $H+L^*,L\%$ differed significantly from $H+!H^*,L\%$ in WAVE.

The SLOPE in $H+!H^*,L\%$ contours was more steeply falling than in $H^*,L\%$, and significantly different from the rising SLOPES of $H^*,H\%$ contours. However, no significant differences emerged between $H+!H^*,L\%$ and $!H^*,L\%$. There was a significant interaction with STRESS: in words with ultimate stress, $H+!H^*,L\%$ contours were more strongly dome-shaped than in words with penultimate stress.

$H^*,L\%$ were distinguished from $H^*,H\%$ contours by their PARABOLA, which was dome-shaped in $H^*,L\%$, but cup-shaped in $H^*,H\%$ contours. $H^*,L\%$ contours differed from their downstepped counterparts only in terms of AVERAGE.

$H^*,H\%$ contours, finally, had a significantly higher AVERAGE than items labeled $!H^*,L\%$, but differed from them also with respect to the SLOPE (rising in $H^*,H\%$ and falling in $!H^*,L\%$).

The preceding analyses have shown that the majority of contours differed significantly from each other in at least one coefficient. However, with this type of analysis (especially when dealing with many categories), the internal structure of the data often remains rather opaque. A technique that can visualize data structures in an intuitive and easy-to-interpret format is classification trees. Classification trees predict the membership of cases (items) in the classes of a categorical dependent variable from their values on one or more predictor variables, and display the outcome of this process in a tree-like format. The structure of the tree reflects the structure of the data in the sense that similar cases end up in the same branch of the tree, and cases that bear less resemblance with each other end up in branches that are further apart. In

8.3 Results

our case, we would like to predict the CONTOUR of an item from the four variables c_0 , c_1 , c_2 and c_3 . To do this, I used the CART (Classification And Regression Tree) algorithm (Breiman, Friedman, Olshen, & Stone, 1984) as implemented in the `rpart` function in R. In a first step, the algorithm looks at all the predictor variables and selects the one that is most useful for splitting the data into two subsets which are each more homogeneous than the original data set. For each of these two subsets, the algorithm then creates two new subsets and so on. The resulting tree for this data set is shown in Figure 25.

8.3 Results

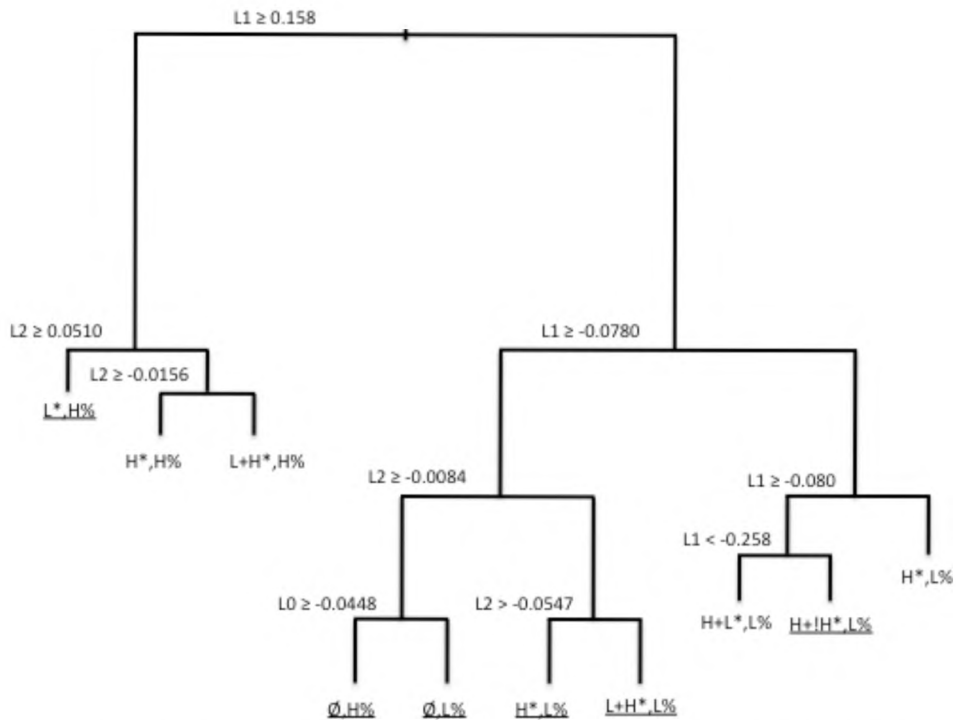


Figure 25: CART tree for CONTOUR as predicted by the four polynomial coefficients (c_0, c_1, c_2, c_3). The expression at each split informs about the decision rule (e.g., “if $c_1 < 0.158$, follow the left branch, otherwise the right” at the first split). Displayed is the initial tree, branches that remained after cost-complexity pruning are set in bold face.

The tree captures the differences in shape between the contours in an elegant way. For example, predominantly rising contours are found in the right part of the tree, whereas contours that are mostly falling are located in the left part. This is achieved by the first splitting criterion, which is SLOPE: Contours with high positive c_1 (i.e., a predominantly rising slope) are sent to the right, contours with a mostly negative c_1 to the left. Another illustrative example is how PARABOLA (c_2) is used to sort $H^*,L\%$ and $L+H^*,L\%$ into different leaves: A

8.4 Extending the model: Adding alignment parameters

given contour is more likely to be L+H*,L% if its c_2 is smaller than -0.054 (i.e., if it is more strongly dome-shaped). Note at this point that this does not necessarily mean that the difference between the two contours in this coefficient is statistically significant; the criterion informs us that for the algorithm, the “best bet” for CONTOUR is L+H*,L%, if the contour has a c_2 smaller than -0.054.

In Appendix C of their article, Grabe et al. (2007, p. 305ff.) show in a constructed example how differences in the coefficients reflect differences in fine phonetic detail of the contours, such as alignment of F0 peaks with the segmental string. They demonstrate that a change in alignment will always result in modifications in SLOPE and PARABOLA. We have also seen this effect in the present data: The contours L+H*,H% and L*+H,H% (which should differ in the position of the pitch minimum) differed in PARABOLA. However, while the general physical interpretation of the first four coefficients is straightforward, “translating” their values into alignment information is perhaps less obvious. This led me to investigate a different way of ascertaining alignment information, which is more easily interpretable to researchers in the AM framework. In the next section, I will describe an extension to the polynomial model, showing how relative alignment parameters can be derived from the set of coefficients that describe a contour.

8.4 *Extending the model: Adding alignment parameters*

Each contour in the data is described by a polynomial function (cf. Equation 1). From this function, the position of local minima and maxima of the curve can be derived. In the present study, this was done by first using the polynomial functions calculated by Polyfit to create polynomial objects in Praat, and then applying customized Praat scripts to compute the local minima and maxima of these objects. Due to the standardization of the scale, this position will be between -1 and 1. For example, in the left panel of Figure 23 above, the F0 peak is located approximately around -0.5. By measuring the absolute position and duration of the lexically stressed syllable and mapping it onto the normalized time scale, it is possible to define the location of the turning points relative to the syllable structure. Here, the location is expressed in percentage

8.4 Extending the model: Adding alignment parameters

of the syllable's duration (e.g., the maximum may occur at 45% into the lexically stressed syllable).

I calculated the alignment of the F0 minimum in the nuclear contours L+H*,H%, L*+H,H% and L*,H%, and tested the differences between the three nuclear contours statistically, using LME models again, with the position of the F0 minimum (POSMIN) as dependent variable, CONTOUR as independent variable, and ITEM and SUBJECT as crossed random factors, where appropriate (as determined by likelihood ratio tests). Because we have seen that the position of the lexically stressed syllable affects the shape of L+H*,H% and L*+H,H% contours, separate analyses were performed for ultimate and penultimate stress words.

For penultimate stress words, CONTOUR turned out to be a significant predictor for POSMIN: The F0 minimum was aligned significantly later in L*+H,H% nuclear contours compared to L+H*,H% nuclear contours (5.9% into the stressed syllable vs. -21.8% preceding the stressed syllable; $p < .001$), and the F0 minimum in L*,H% nuclear contours was in turn aligned significantly later than that of L*+H,H% nuclear contours (39.7% into the stressed syllable vs. 5.9% into the stressed syllable; $p < .001$). The differences in alignment are illustrated Figure 26.

8.4 Extending the model: Adding alignment parameters

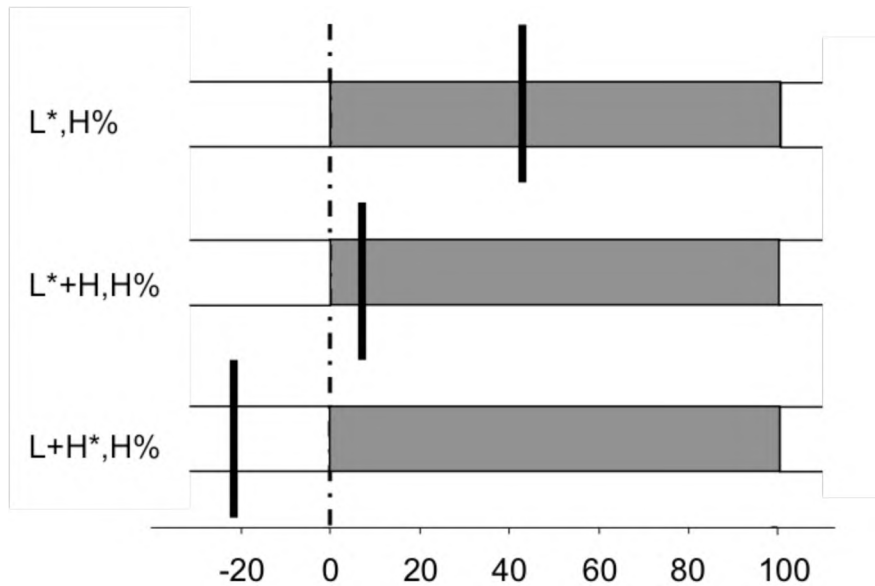


Figure 26: Graphical representation of the relative position of the F0 minimum (vertical black lines) for the three nuclear contours L+H*,H%, L*+H,H% and L*,H% (for penultimate stress words), expressed as percentage of the stressed syllable duration (x-axis). The grey-shaded area marks the lexically stressed syllable, the vertical dotted line indicates the syllable onset.

For words with ultimate stress, CONTOUR was a significant predictor for POSMIN as well. The F0 minimum was aligned significantly later in L*+H,H% nuclear contours compared to L+H*,H% nuclear contours (13.6% into the stressed syllable vs. -2.3% preceding the stressed syllable; $p < .05$). Furthermore, the F0 minimum in L*,H% nuclear contours was aligned later than that of L*+H,H% nuclear contours (23.1% into the stressed syllable vs. 13.6% into the stressed syllable), yet the difference did not reach statistical significance ($p = .1$).

These results are in line with the descriptions of the accents in the GToBI literature (Benzmüller & Grice, 1997; Grice & Baumann, 2002; Grice et al., 2005): In L+H* accents, the F0 minimum occurs in the pre-tonic syllable. In L*+H accents, the trough is usually located in the stressed syllable, but may be

8.4 Extending the model: Adding alignment parameters

pushed to the left if there is tonal crowding. Tonal crowding occurs when several tones are associated with the same tone-bearing unit. This can lead to small differences in alignment (Arvaniti, Ladd, & Mennen, 2006a, p. 670). In the case of L*+H,H% contours in this data set, both the high trailing tone of L*+H and the high boundary tone (H%) are aligned with the last syllable in the phrase (in penultimate stress words). As a consequence, the F0 minimum occurs earlier, at the beginning of the stressed syllable. The F0 minimum in the L* accent, finally, should be located clearly within the stressed syllable. This is the case in this data as well.⁷¹

Thus, relative alignment parameters derived from polynomially modeled curves are able to capture the phonetic differences between contours in an easily interpretable format. This way of representing alignment is perhaps more useful with more variable speech material than the plain coefficients. Finally, deriving the F0 turning points from the modeled curve is a less subjective procedure than determining them manually from the original pitch track. As discussed in chapter 1, tonal targets cannot always be unambiguously identified because of discontinuities in the visible contour, which means for the labelers that they have to make (sometimes arbitrary) decisions as to where to locate the targets. This is not the case with this procedure.

Before moving on to the general discussion, I will present another potential use for polynomial modeling of pitch contours. With its capacity to capture fine phonetic detail, the polynomial approach allows not only for “comparisons of intonational systems across dialects and languages” (Grabe et al., 2007, p. 298), it may also provide a window into developmental changes in pitch contour realization. In the next section, I will look at how the same pitch contours have been realized by the three different age groups.

⁷¹ For ultimate stress words, the difference in alignment between L*+H,H% contours and L*,H% contours is smaller than for penultimate stress words. This could be explained by the fact that in ultimate stress words, all tones of the contour are associated with the final syllable. In both contours, tonal repulsion pushes the F0 minimum towards the left edge of the final syllable.

8.5 Investigating age-related differences with polynomials

8.5 *Investigating age-related differences with polynomials*

The field of child intonation research has become increasingly popular over the past years (for a review, see Snow & Balog, 2002). However, the development of pitch accent realization has not received much attention. As already mentioned in the preceding chapters, several imitation studies report that children up to the age of seven may have difficulties in producing some F0 patterns in an adult-like way (Loeb & Allen, 1993; Patel & Grigos, 2006; Snow, 1998), but these studies were more concerned with global pitch trends rather than specific pitch accents or contours. The phonetic analyses in chapter 4 produced evidence that there appear to be differences in the way children and adults realize certain intonation contours. We would therefore also expect to find differences between the age groups in one or more of the coefficients that describe their contours.

There is, to my knowledge, only one study that has used polynomial modeling to investigate pitch accent realization in children, but it used a different model and had a different objective. Ota (2008) analyzed the productions of Swedish toddlers in order to find out whether they reliably produce the pitch contour of Accent II.⁷² Thus Ota was not interested in a developmental trend, but in whether the children were able to produce a specific contrast (here: distinguishing disyllabic Accent II words from other disyllabic productions). His analysis involved the stylization of the children's F0 contours using the MOMEL algorithm (Hirst et al., 1993). MOMEL looks for a continuous series of quadratic second-degree non-orthogonal polynomials that offer the best fit to the visible F0 curve, and marks the inflections as turning points. To determine whether the children realized the Accent II contour correctly, Ota checked whether the children's productions contained the sequences of high and low turning points (as identified by the algorithm) that were appropriate for the accent. Indeed, he found that the children produced the F0 pattern typical for Accent II words, indicating that they have acquired this phonological category. Ota's study did not compare the children's production to that of adult speak-

⁷² Swedish has two lexically contrasting pitch patterns associated with the stressed syllable, which are usually referred to as Accent I and Accent II.

8.5 Investigating age-related differences with polynomials

ers, and the author himself raises the possibility that the children's phonetic realization of Accent II contours differs from adult realizations (Ota, 2008, p. 244). In the present study, I am particularly interested in this potential difference. We will look at how children may differ from adults in the specific realization of several nuclear contours, that is, if they show differences in the coefficients that describe their contours. Polynomial descriptions offer an ideal way for studying phonetic differences *within* phonologically defined contour types, as differences in the phonetic realizations of one accent category will be reflected in differences in the coefficients. Note that because MOMEL describes contours in a *series* of polynomial expressions rather than in a single one like Polyfit, it would be difficult to compare productions with each other. In the MOMEL-approach, some F0 contours may be described by a series of two polynomial expressions, while others may be described by three. It is unclear how one could determine the degree of (non-)similarity of the contours statistically. In addition, non-orthogonal polynomials (as the ones implemented in MOMEL) have the disadvantage that their coefficients are not independent of each other and are often highly correlated, which would have to be taken into account in the statistical analyses (Grabe et al., 2007, p. 287). In contrast, Polyfit-modeled contours are described in one set of minimally correlated Legendre polynomials, which can be straightforwardly subjected to analyses between age groups. This is exemplified in the following paragraphs.

Table 23 shows how often each age group produced each nuclear contour. The distributions are uneven, with some nuclear contours used much more often by one age group than by the others, and some contours hardly or not at all produced by one age group. Therefore I only looked at those nuclear contours of which there are at least seven tokens in each age group (the grey-shaded rows in Table 23). These were the same contours that were also investigated in chapter 4, section 4.3.

8.5 Investigating age-related differences with polynomials

Contour	5 years	7 years	adults
∅,L%	15	9	5
∅,H%	1	9	6
L+H*,L%	8	9	7
L+H*,H%	4	8	11
L*,H%	5	10	11
L*+H,H%	7	18	22
H+L*,L%	1	0	9
H+!H*,L%	14	12	8
H*,L%	8	25	7
H*,H%	4	9	6
!H*L%	9	5	2
other	1	3	3

Table 23: Frequency of the different contour types by age. Nuclear contours that were produced at least seven times in each age group are shaded in grey.

For each of the four contours (L+H*,L%, L*+H,H%, H+!H*,L% and H*,L%), I initially specified a LME model for each of the four coefficients (c_0 , c_1 , c_2 , c_3), with the coefficient in question as the dependent variable and age (AGE) as fixed factor, and SUBJECT and ITEM as random factors. Since the variance explained by these factors was in most cases so low that a simple linear model without random factors (the LM function in R) was fitted instead. Due to the small sample sizes, the statistical power is rather low. For this reason, p-values smaller than .05 will be regarded as significant, and p-values up to and including .07 will be reported as marginally significant. There were significant or marginal effects of AGE for at least one coefficient in three of the contours. No significant differences were found for H*,L%. The results are summarized in Table 24.

8.5 Investigating age-related differences with polynomials

Contour	Coeff.	effect of AGE	p
L+H*,L%	c ₁ (SLOPE)	5 years (0.02) = 7 years (-0.03) < adults (0.16)	< .05
L+H*,L%	c ₃ (WAVE)	adults (-0.09 ⁷³) > 7 yrs. (0.01) = 5 yrs. (0.00)	= .06
L*+H,H%	c ₀ (AVER.)	adults (0.15) < 5 years (0.42)	= .05
L*+H,H%		7 years (0.09) < 5 years (0.42)	< .05
L*+H,H%	c ₁ (SLOPE)	adults (0.44) < 5 years (0.61)	= .07
L*+H,H%		7 years (0.41) < 5 years	< .05
H+!H*,L%	c ₃ (WAVE)	adults (-0.09) > 7 yrs. (0.00) = 5 yrs. (0.00)	< .05

Table 24: Significant and marginally significant effects of age on coefficients for three nuclear contours. The contour concerned is listed in the left-most column, the second column gives the coefficient for which a significant effect of AGE was found, the third column informs about the direction of the effect, with the coefficient estimates in brackets. The right-most column gives the probability. If the probability for two comparisons (e.g., adults – 7 years and adults – 5 years) had the same level, the results are summarized in one row.

The negative second coefficient for seven-year-olds in L+H*,L% indicates that the contour had a larger falling portion than in adults, where the SLOPE was predominantly rising. The fact that the fourth coefficient (WAVE) was near zero in both child groups suggests that their realization of this contour was much less strongly falling-rising-falling, or less “wiggly” than that of adults. In L*+H,H% contours, five-year-olds differed from both seven-year-olds and adults in that they had a higher first coefficient (AVERAGE). This means that they produced these contours higher up in their pitch range than the two other age groups. At the same time, five-year-olds produced L*+H,H% contours also with a steeper rise than seven-year-olds and adults. Finally, both five-year-olds and seven-year-olds differed significantly from adults in the fourth coefficient of H+!H*,L%. While the adults’ negative c₃ shows a falling-rising-falling movement, the children’s near-zero c₃ indicates an absence of

⁷³ The vertical lines indicate that although the adults’ coefficients are smaller in relative terms, the “absolute movement” is larger.

that wave shape. The results of the by-age analysis will be discussed along with the other results in the next section.

8.6 *General discussion*

I modeled the F0 of spontaneously produced hand-labeled German nuclear contours quantitatively in terms of orthogonal Legendre polynomials. Statistical analyses showed that the majority of these nuclear contours differ from each other significantly in at least one coefficient. This lends support to the accuracy of the assigned GToBI labels. In what follows, I will discuss some selected examples. As one would expect, all of the rising contours such as L*,H% and L*+H,H% were found to be different from the falling contours such as H*,L% and H+!H*,L%, either in terms of SLOPE (predominantly rising vs. predominantly falling) or in terms of PARABOLA (more cup-shaped vs. more dome-shaped). However, the coefficients also picked up on differences within the group of rising and falling contours. For example, I found that the contour H+L*,L% had a more steeply falling slope than H*,L%. Within the group of rising contours, L*,H% proved different from L*+H,H% contours for instance in the cup-shaped PARABOLA (the third coefficient), which was “narrower” in L*,H% contours. Importantly, the analysis of my data showed that polynomial modeling was not only able to distinguish between different nuclear contours, but also (with two exceptions, see below) between accented contours and deaccented contours. Their averages differed significantly from each other in the first coefficient (AVERAGE) and in the second coefficient (SLOPE).

But there were also four statistically non-significant differences. Three of these involved H*,L% or its downstepped counterpart !H*,L%. The accents H* and !H* have the same shape, but differ in their relationship with the preceding pitch peak. While the pitch maximum of H* accents is typically about the same height as that of the preceding high pitch accent or slightly lower, !H* accents are clearly lower than the preceding pitch peak. Hence, in order to be able to distinguish H* accents from downstepped accents (!H*), information about the previous high pitch accent is necessary. However, the analyses showed that the two contours differed from each other, namely in their mean F0. This is not surprising, given that !H* accents especially in phrase-final posi-

8.6 General discussion

tion would be expected to be lower in the speaker's pitch range than H* accents. At the same time, the fact that the two contours share the same overall shape justifies combining them in one category, and this is also what Grabe et al. (2007) did in their analyses. Against this background, I collapsed the two contour categories in the present data (resulting in 55 items in the new H*,L% category) and re-ran the analyses. As a result, the previously non-significant differences with the other contours became statistically significant: the difference between H*,L% and H+!H*,L% was significant (for c_0 and c_1 $p < .05$), and so was the difference between H*,L% and \emptyset ,L% (for c_2 $p < .01$). The difference between H*,L% and L+H*,L% was marginally significant for SLOPE ($p = .05$).

These results show that the contours that were assigned different phonological labels clearly differed in the average shape of their F0 contours, paralleling the findings by Grabe et al. (2007) for English read-out speech. Unlike the Grabe et al. corpus, the German data analyzed here also contained deaccented items followed by high and low boundary tones. These proved to be significantly different from accented items, mostly in terms of either SLOPE or PARABOLA, in all except one case (\emptyset ,H% vs. H*,H%). This is evidence that the difference between accented and deaccented words is typically reflected in the polynomial coefficients that describe their F0 contours. A potential explanation for the fact that there was no significant difference between deaccented items and H* before high boundary tones is that this contrast may be signaled by other acoustic parameters such as intensity and duration, which have been found to be correlates of prominence (e.g., Kochanski et al., 2005).

The comparison between H*,L% and !H*,L% has shown that even downstep can be captured with polynomial modeling, as downstepped contours are typically produced lower in the speaker's pitch range (i.e., with a lower first coefficient) than non-downstepped contours. However, especially when dealing with small to medium-sized data sets, it seems useful to combine downstepped and non-downstepped contours in order to gain more statistical power for comparisons with other contour types. Like in the English data, models using the first three coefficients ($c_0 - c_2$) were sufficient to distinguish the majority of nuclear contours in this data set from each other. The contour pair H+!H*,L% and H+L*,L%, though, differed from

each other only in the fourth coefficient, WAVE, suggesting that this coefficient is necessary to distinguish these two contours from each other. This is somewhat surprising, as one would probably also expect a difference in SLOPE (see Table 17). In fact, we have seen that SLOPE was the most useful splitting criterion for the CART algorithm (see Figure 25).

The fact that the difference between these two contours was small could be seen as support for the proposal that they are actually phonetic variations of the same accent type (e.g., Grice, Baumann, & Jagdfeld, 2007; Rathcke & Harrington, in press). However, it needs to be borne in mind that H+L*,L% contours occurred only ten times in the data. More data are needed to find out whether the two contours differ consistently from each other in their coefficients.

To what extent statistically significant differences in coefficients correspond to differences in the perception of human listeners is yet another question (Grabe et al., 2007, p. 296). Perception experiments that systematically vary the coefficients that describe an F0 contour can shed more light on this issue. For these types of experiments it is advantageous that polynomial expressions can serve directly as input for speech synthesis.

In addition to applying the polynomial approach to German data, this study has introduced a way to derive relative alignment measures from the polynomial models, which are more easily interpretable than the plain coefficients. The position of tonal targets is expressed relative to the duration of the lexically stressed syllable. Using this parameter, the three rising contours L+H*,H%, and L*+H,H% and L*,H% were shown to differ in their alignment of the F0 minimum in the manner that is described in the GToBI literature (Benzmüller & Grice, 1997; Grice & Baumann, 2002; Grice et al., 2005). Baumann (p.c.) argues that it would not be possible to distinguish L*+H,H% and L*,H% contours from each other in terms of alignment of the pitch minimum, as the trailing tone of the L*+H accent would be merged with the high boundary tone. This would mean that the contour would be identical with L*,H%. However, the results suggest that there are two types of contours in the data, one in which the pitch minimum occurs at the beginning of the stressed syllable, and one in which it occurs more towards the middle of the stressed syllable.

8.6 General discussion

ble.⁷⁴ The alignment parameter provides intonational phonologists with a simple tool to compare how accents differ in the alignment of their tonal targets with the segmental string, avoiding the problems associated with manual location of pitch events. In this way, the polynomial approach can be smoothly integrated with the productive line of research on “segmental anchoring” (among others, Arvaniti et al., 1998; Arvaniti, Ladd, & Mennen, 2006b for Greek; Atterer & Ladd, 2004 for German; Prieto, Santen, & Hirschberg, 1995 for Spanish; Silverman & Pierrehumbert, 1990 for American English).

Finally, I have shown how polynomial modeling can be used to investigate differences in pitch contour realization across age groups. Although the sample sizes are small, the within-category analysis of four nuclear contours has yielded some interesting findings. For instance, I found that in two contours, H+!H*,L% and L+H*,L%, both the five- and the seven-year-olds hardly varied the fourth coefficient, WAVE. This is a sign that their contours were not as “wiggly” as the ones produced by the adults. As explained before, the higher-ranking coefficients pick up the more rapidly changing properties of the F0. The fact that the children’s c_3 was so small may be an indication that they are not yet able to modulate their pitch as fast as adults. Previous studies have reported that children may experience difficulties with the adult-like production of rising contours (Loeb & Allen, 1993; Patel & Grigos, 2006; Snow, 1998). However, these studies have concentrated on either the falling portion or the rising portion of an contour, not on the overall movement of a contour, such as the fall-rise-fall movement in L+H*,L% contours. This movement is reflected in the fourth coefficient of the polynomial model. Thus polynomial modeling may enable us to gain even better insights into children’s pitch accent realizations. More research (for example along the lines of Xu and Sun, 2002) is needed to obtain more conclusive information about the developmental trajectory of speed of pitch change control.

⁷⁴ Note that it is not the case that there are more *Kamel* items in the L*,H% category. A Fisher’s exact test showed that contour category (L*+H,H% vs. L*,H%) is independent of rhythmic meter (iamb vs. trochee), $p = .6$.

With regard to the age group differences found here the same caveat applies as with the differences between contour types: The differences in coefficients may not be perceptible. On the other hand, they can still be informative. In connection with the acquisition of segmental phonology the term “covert contrasts” (Scobbie, Gibbon, Hardcastle, & Fletcher, 1996) has been used to describe situations in which children already produce a contrast between two categories (e.g., the voice onset time contrast between voiced and voiceless stops such as /t/ and /d/), but because the realizations fall within one perceptual category for adults, the contrast is not recognized by adult listeners. Analyzing children’s speech acoustically can recover these imperceptible differences and can tell us something about what a speaker has already acquired (Scobbie et al., 1996, p. 44). The same rationale can be applied the other way round: Even though children may produce sound patterns (such as accents) that are perceptually equivalent to adult productions, there may still be systematic underlying differences, which can tell us something about the way the speech patterns mature with age.

This study has shown that the polynomial approach can also be applied to German natural speech, and the results lend additional support to the accuracy of the labels assigned in chapter 4. However, a final remark about the possibilities and limitations of this method is in order. We have seen that the polynomial approach may be a useful way to check if – within a restricted domain – contours that a human labeler judged to be different are also different in terms of a more objective mathematical description. In this sense, the use of the technique is comparable to that of algorithms for elbow location, that is, as a tool to reduce subjectivity in prosodic analysis. In addition, as shown in this chapter, the approach has potential merits for the study of within-contour variation across speaker groups. However, because polynomial modeling is concerned with the optimal approximation of *surface* realizations, it cannot capture the underlying linguistic or functional equivalence of two contours that are realized on domains of differing length. For instance, polynomial descriptions would probably fail to grasp the similarity between Ladd’s (2008) two ‘disbelief’ contours presented in the background (*Sue!?* and *A driving instructor!?*). This can be captured only with recourse to an abstract

8.6 General discussion

phonological level (see Arvaniti & Ladd, 2009 for an excellent discussion of this issue). Therefore polynomial models can never be a substitute for phonological theory.

9 Summary and conclusions

This dissertation has investigated how German children and adults use intonation to mark the topic-hood and the information status of referents. In addition, some general methodological issues concerning the analysis of child and adult prosody were investigated.

Previous research on information structure and intonation in German has been based either on reading (with adults) or on imitation tasks (with children). In contrast, the studies in this dissertation were based on a corpus of unconstrained speech, elicited by means of a picture story telling task. The present investigation differed also from other studies in the field of child intonation, in that it combined phonological and phonetic analyses. This last chapter summarizes the results of the studies presented in chapters 3 to 8, and concludes with suggestions for future research.

The first study in chapter 3 looked at the prosodic realization of topical referents in the so-called *pre-field* (i.e., the position preceding the finite verb). The results showed that adults realized topical referents predominantly with the rising accents, L+H* and L*+H, in line with previous findings obtained from read speech (Braun, 2006; Mehlhorn, 2001). However, I also found a considerable number of H* accents and deaccented referents, which have so far not been attested in experimental studies in German. The five- and seven-year-old children used the same set of pitch accents as the adult group, but differed in the frequency with which they used the different accent types. They used more H* accents than the adult group, and the youngest speakers hardly produced any L*+H accents. Some differences between child and adult speakers were also found at the phonetic level. Compared to adult speakers, seven-year-olds' accents had a smaller excursion, and children of both age groups produced accents with a slower speed of pitch change (shallower slope) than the adults. Furthermore, they aligned pitch minima earlier than the adults, while the location of the pitch maxima was adult-like. A possible explanation for the earlier alignment is that children initiated the rising movement earlier to reach the high target in order to compensate for the slower speed of their

9 Summary and conclusions

pitch change. Still, the study showed that children use the same accent types as adults and do not mark topical referents in a fundamentally different way.

However, a clear difference between adults and children was observed with respect to the sheer number of topic referents in the pre-field. Referents in sentence-initial topic position were generally hard to find in the child data. Unlike the adults, children filled the pre-field often with connectives and temporal adverbs, such as *und dann* 'and then', an observation that has been reported in other studies on child narratives as well (Hickmann et al., 1996). While this finding is not directly pertinent to the research question at hand, it reveals something about different strategies of structuring discourse across ages. Adults assume that – unless indicated otherwise – time progresses sequentially in narratives, and they do not necessarily mark this lexically. They pay more attention to the global thematic structure of a story and build their narrations around the protagonists and their actions. Children, on the other hand, seem more concerned with the local temporal linking of events, which they express by these connective and temporal adverb constructions. As discussed below, the data on the realization of phrase-final referents (chapter 4) suggests that while children link local events temporarily using lexical means, adults do this less frequently and rather use intonation to signal that they are not yet finished with their narration.

Chapter 4 investigated the marking of information status of discourse referents, more specifically, the marking of activation. The discourse referents in the picture stories had one of three information states: new, given, or accessible. Speakers always accented new referents, but there was no evidence for a designated newness accent. An unexpected finding was that L* and L*+H accents were used more often for new referents than for given referents. This is surprising, as accents containing L* have been suggested to be markers of givenness (Pierrehumbert & Hirschberg, 1990). The present results clearly do not support this claim. Because speakers generally seem to have a preference to combine contrasting pitch accents and boundary tones, that is, combining low pitch accents with high boundary tones and vice versa (Dainora, 2002b), it was hypothesized that the high occurrence of L* containing accents was a consequence of the wide-spread use of high boundary tones to indicate continua-

tion. The results show that some generalizations (e.g., 'new referents are marked with H*') might not hold when intonation is looked at in a less restricted context; in this case, in a context including high and low boundary tones. In contrast to new referents, referents in the given condition were more often deaccented, in line with previous reports (e.g., Baumann, 2006; Baumann & Hadelich, 2003; Brown, 1983), but the fact that more than half of all given referents were still accented shows that this is only a tendency. There was no evidence for a special accessibility accent H+L*, as suggested by Baumann and Grice (2006). Accessible referents patterned with new referents in terms of pitch accent type, but ranked between given and new referents regarding the proportion of deaccented referents. This finding supports Lambrecht's (1994) claim that there is no clear phonological correlate of accessibility, but that it is up to the speaker whether to treat accessible referents like given or like new ones. On the whole, these results call into question a direct mapping between information status and pitch accent type as has been proposed in the literature (Baumann, 2006; Baumann & Grice, 2006; Chen et al., 2007; Pierrehumbert & Hirschberg, 1990). Rather, the findings support the basic assumption that accentuation of new referents is obligatory, whereas the realization of given (and accessible) referents is less constrained.

With respect to language development, the results show that children as young as five years old have learned to use intonation to mark information status in an adult-like fashion. They signal newness by accenting, and know that givenness (previous mention) licenses deaccentuation. This is in accordance with what has been reported for simple sentence pairs (Wonnacott & Watson, 2008), but the results of the present study establish that this also extends to more complex discourse structures such as narrations. The fact that the children treated given and accessible referents differently furthermore demonstrated that – like adults – the children were sensitive to recency of mention, and used intonation to indicate this. This early intonational competence in information status marking contrasts with the difficulty children of the same age have in information status marking in the morphosyntactic domain (e.g., Kail & Hickmann, 1992). The results also show that already at the

9 Summary and conclusions

age of five, children have acquired a large part of the adult pitch accent inventory.

At the same time, the study also revealed some age-related differences. Compared to the seven-year-olds and the adults, the five-year-olds produced only few utterances with phrase-final rising intonation, a pattern that can serve as a signal to the listener that the speaker is about to say more. It was hypothesized that children may acquire this discourse-structuring (or interactional) function of intonation only later. This is particularly interesting in the light of the finding from chapter 3 that the children used lexical means (like *und dann*) to indicate temporal continuity. It is conceivable that children start out with a lexical strategy to achieve cohesion, and later move on to use continuation intonation for this.

Another age difference emerged for the accent H+L*, which was essentially absent in both child groups. Drawing on experimental findings by Dombrowski (2003) and Kohler (1991) that H+L* can convey an attitude of “knowingness”, it was hypothesized that adult speakers may have used this accent in this function, and that paralinguistic uses of intonation to express attitudinal meanings may be another aspect which is acquired only later. Alternatively, H+L* may be part of a certain register used for story telling, which is only employed by adult speakers.

In addition to the phonological analyses, chapter 4 also analyzed how the speakers of the different age groups realized the phrase-final accents phonetically. The results showed that children produced the rising portion of accents with a smaller excursion and shallower slopes than adults. The findings mirror the results for sentence-initial (pre-nuclear) accents, and suggest that although children differentiate between a number of accent types in the same way as adults do, their realization of individual accents may still not be fully adult-like by age seven.

The last chapter of part II, chapter 5, examined children’s ability to differentiate identifiability from activation. For this, I analyzed children’s narrations of stories in which well-known characters, *der Sandmann* ‘the sandman’ and *der Weihnachtsmann* ‘Santa Claus’, were introduced. These referents were identifiable for both speaker and hearer due to cultural common ground, but

new to the discourse. Like the adult group, children used the same pitch accent types to realize these unused referents as they had used for the brand-new referents in chapter 4. This indicates that they have learned that the identifiability of a referent due to cultural common ground does not influence the way the referent's activation is marked locally.

Taken together, the results of chapters 3 to 5 suggest that five- and seven-year-old children have acquired most of the adult pitch accent type inventory, and know how to use these forms appropriately to mark information-structural categories. Where they diverge from adults is the use of intonation for discourse-structuring and paralinguistic purposes on the one hand, and the exact phonetic realization of pitch accents on the other.

In part III of the dissertation, I investigated some central methodological issues. In the study in chapter 6, I tested whether the discrepancies between the findings obtained in chapter 4 and the reports in the literature were due to differences in speaking mode: The data analyzed in chapter 4 came from spontaneously produced narrations, whereas previous studies were based on scripted speech. To answer this question, a group of adult speakers read out texts, which described and accompanied the depicted scenes from the picture stories used in chapter 4. In terms of information status marking, the results replicated earlier findings from Baumann's (2006) corpus study of read-out newspaper texts. This sheds new light on the unexpected results from chapter 4: Intonation in spontaneous speech is different from intonation in reading, and the same results as Baumann's are obtained when speakers are given a reading task. Specifically, speakers always deaccented given referents, and marked accessible referents preferably with H+L*. Moreover, H+L* accents were used frequently in general, paralleling results from other reading studies (Braun, 2006). The experiment shows that there are clear differences in the intonation patterns of read speech and spontaneous speech. It is argued in chapter 6 that this might be because the various functions of intonation receive different weights, depending on the communicative situation. In spontaneous speech, floor-holding or signaling of attitudes can be more important than the marking of information status. Conversely, in reading, many paralinguistic functions play only a minor role, allowing speakers to concentrate on

9 Summary and conclusions

using intonation to fully reflect the information structure of the text. I furthermore hypothesized that the typical intonational marking of information structure that is postulated in the literature may be a cultural skill that is learned in the course of literacy training. The findings from chapter 6 have important implications for first language acquisition research. They show that we cannot assume intonation in scripted speech to be the target model for children. Moreover, they underscore the necessity to test claims based on constrained speech also with more naturalistic data.

The experiment reported in chapter 7 investigated naïve listeners' perception of prosodic boundaries in the corpus. The study was motivated by two factors. First, when annotating the data for the analysis of phrase-final referents in chapter 4, it had at times been difficult to decide whether a target referent was followed by an (IP) boundary, or not. If untrained listeners agreed on the location of phrase boundaries, their judgments could be used to inform and legitimate these decisions. Second, the existing evidence in the literature is inconclusive with respect to children's ability to mark prosodic boundaries. The level of agreement among naïve raters could serve as an indicator of the reliability of boundary cues that the children produce. In the experiment in chapter 7, adult participants listened to utterances from the corpus in real time, and indicated on a written transcript where they perceived phrase breaks. The results showed substantial overall agreement among the raters, comparable to that obtained in other studies (e.g., Mo et al., 2008). Importantly, the agreement scores were similar across age groups. This finding suggests that the method is a suitable supplementary tool for the analysis of spontaneous child speech. In addition, the results provide evidence that the children produced clear acoustic cues to signal prosodic structure.

Finally, chapter 8 turned again to the issue of finding quantifiable evidence for prosodic labels. The outcome of the checking procedure used in chapter 4 suggested that the data had been labeled consistently. However, the procedure is time-consuming, and the determination of acoustic landmarks like F0 turning points is not always straightforward and sometimes susceptible to subjective bias. In chapter 8, an alternative method was explored: Nuclear contours (i.e., nuclear accents and the subsequent boundary tones) were

modeled mathematically using third-order Legendre polynomials, following an approach similar to the one Grabe, Kochanski and Coleman (2007) used for English. Statistical analyses showed that when H*,L% contours and !H*,L% contours were combined in one category, all but one contour pair differed significantly from each other in at least one of the four coefficients. These results are in accordance with those obtained with the other procedure, and lend additional support to the accuracy of the assigned labels. Furthermore, how information about the alignment of tonal targets relative to the syllable structure can be derived from the polynomial descriptions was demonstrated, making the approach compatible with more linguistically oriented labeling schemes. Finally, a within-accent category comparison across age groups revealed interesting differences between child and adult speakers. Findings such as a lower second coefficient (slope) in children replicate what had been found in the manual phonetic analysis in chapter 4. Other differences between age groups were only obtained with the polynomial method, such as a lower third coefficient in the children's accents, which could be a sign of a lower control of dynamic pitch in children.

Taken together, the methodological studies presented in chapters 5 to 8 have yielded two main results: First, intonation from scripted speech cannot readily be adopted as the target model for children. In order to know what children strive to attain, it is also necessary to obtain information about what they are exposed to in more natural settings. Second, researchers working on child prosody can avail themselves of methods that go beyond purely impressionistic labeling or aggregate measures like mean pitch. Naïve listener judgments can inform expert labeling, and mathematical modeling can provide quantifiable acoustic evidence to support phonological analyses, as well as offer new insights into speech development.

This dissertation has produced several results that invite further investigation. Overall, the results indicate continuity in intonational marking across the age groups. We have seen many similarities between adults and children, showing that five-year-olds are already very apt users of intonation to structure their utterances. However, differences emerged on two levels: First, the acoustic analyses suggest that the phonetic 'fine-tuning' of pitch accent reali-

9 Summary and conclusions

zation is still ongoing even in seven-year-olds. Whilst this does not seem to impede children's ability to mark information structure, it would be interesting to know when their production becomes fully adult-like. In this connection the present data also begs the question of whether the relative scarcity of L*+H accents that was observed in the youngest age group has purely physiological reasons, or whether the children just happened not to use this accent in the contexts that were investigated here. Second, adults and young children seem to differ in the use of prosody for interactional purposes. Unlike the two older age groups, the five-year-olds did not produce many sentence-final rising contours, which can serve as continuation and turn-holding device. They also produced fewer filled pauses to indicate delays. Future studies could focus on when and how children employ prosody as an interactional resource.

Another interesting topic concerns the differences between spontaneous and read speech. The results obtained in this dissertation suggest that information structure marking in spontaneous speech is more varied than in read speech. While it is impressive that children acquire adult-like marking in spite of this variability, it would also be interesting to investigate when they actually learn what appears to be a formal "reading register". I have hypothesized that listener judgments may partly be influenced by prescriptive ideas of what constitutes "good intonation" (cf. chapter 4). If this were the case, one may wonder if children who have not yet learned oral reading would judge certain aspects of intonation different from adults, and whether learning how to read aloud may have an effect on their judgments.

It appears that children receive three different kinds of intonational input. First, they are exposed to motherese, second, they hear normal conversations around them, and finally, they hear the stories that their caretakers read to them. An intriguing subject for future research would be to compare the intonation in these different types of input, and their relative influence on the target intonation model that children are striving to attain.

References

- Agresti, A. (1996). *An introduction to categorical data analysis*. New York: John Wiley.
- Agresti, A. (2002). *Categorical data analysis*. New York: John Wiley.
- Allerton, D. J. (1978). The notion of 'givenness' and its relation to presupposition and theme. *Lingua*, *44*, 133-168.
- Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.
- Arnold, J. E. (2008). THE BACON not the bacon: how children and adults understand accented and unaccented noun phrases. *Cognition*, *108*(1), 69-99.
- Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expressions: Everyone counts. *Journal of Memory and Language*, *56*(4), 521-536.
- 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. (2000). What is a starred tone? Evidence from Greek. In M. Broe & J. B. Pierrehumbert & J. Editors (Eds.), *Papers in laboratory Phonology V* (pp. 119-131). Cambridge: Cambridge University Press.
- Arvaniti, A., Ladd, D. R., & Mennen, I. (2006a). Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions. *Journal of Phonetics*, *48*, 667-696.
- Arvaniti, A., Ladd, D. R., & Mennen, I. (2006b). Tonal association and tonal alignment: Evidence from Greek polar questions and contrastive statements. *Language and Speech*, *49*(4), 421-450.
- Arvaniti, A., & Ladd, R. D. (2009). Greek wh-questions and the phonology of intonation. *Phonology*, *26*, 43-74.
- 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.
- Baayen, H. R. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

References

- Bamberg, M. G. W. (1987). *The Acquisition of Narratives - Learning to Use Language* (Vol. 49). Berlin; New York; Amsterdam: Mouton de Gruyter.
- Barr, D. J., & Keysar, B. (2005). Making sense of how we make sense: The paradox of egocentrism in language use. In H. L. Colston & A. N. Katz (Eds.), *Figurative language comprehension: Social and cultural influences*. Mahwah, NJ: Erlbaum.
- Bartels, C., & Kingston, J. (1994). Salient pitch cues in the perception of contrastive focus. In *Proceedings of Focus and Natural Language Processing* (pp. ...).
- Bates, D. M., & Sarkar, D. (2007). lme4: Linear mixed-effects models using Eigen and syntax classes. R package version 0.9975-12.
- Baumann, S. (2006). *The Intonation of Givenness - Evidence from German* (Vol. 508). Tübingen: Niemeyer.
- Baumann, S., Becker, J., Grice, M., & Mücke, D. (2007). Tonal and articulatory marking of focus in German. In *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)* (pp. 1029-1032), Saarbrücken.
- 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. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)* (pp. 1811-1814), Barcelona.
- Beckman, M., & Ayers, E. (1997). Guidelines for ToBI labelling, version 3. Ohio State University: The Ohio State University Research Foundation.
- Beckman, M., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255-309.
- Bennett-Kastor, T. (1983). Noun phrases and coherence in child narratives. *Journal of Child Language*, 10(1), 135-149.
- Benzmüller, R., & Grice, M. (1997). Trainingsmaterialien zur Etikettierung deutscher Intonation mit GToBI. *Phonus*, 3, 9-34.
- Berman, R. A. (1996). Form and function in developing narrative abilities. In D. I. Slobin & J. Gerhardt & A. Kyratyis & J. Guo (Eds.), *Social interaction, social context and language* (pp. 343-368). Mahwah, NJ: Erlbaum.

References

- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech, 38*, 365-391.
- Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication, 14*, 359-375.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Institute of Phonetic Sciences, University of Amsterdam Proceedings, 17*, 97-110.
- Boersma, P., & Weenink, D. (1992-2008). Praat - doing phonetics by computer (Version 5.0.35). Amsterdam: Institute of Phonetics.
- Bolinger, D. (1972). Accent Is Predictable (If You're a Mind-Reader). *Language, 48*(3), 633-644.
- Boutsen, F. R., & Hood, S. B. (1996, Jun 05-08). Determinants of speech rate and fluency in fast and slow speaking normally fluent children. In *Proceedings of the 3rd International Conference on Speech Motor Production and Fluency Disorders* (pp. 557-564), Nijmegen, Netherlands.
- Braun, B. (2005). *Production and Perception of Thematic Contrast in German* (Vol. 9). Frankfurt am Main: Peter Lang Publishing.
- Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Language and Speech, 49*(4), 451-493.
- Braun, B., & Chen, A. (submitted). Intonation of 'now' in resolving scope ambiguity in English and Dutch.
- Braun, B., Weber, A., & Crocker, M. W. (2004). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech, 49*(3), 367-392.
- Brazil, D. (1975). *Discourse Intonation* (Vol. 1). Birmingham: Birmingham University.
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees*. Belmont, CA: Wadsworth.
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: Some uses, misues, and alternatives. *Educational and Psychological Measurement, 41*, 687-699.

References

- Brown, G. (1983). Prosodic Structure and the Given/New Distinction. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 67-77). Berlin: Springer.
- Brown, G., Currie, K. L., & Kenworthy, J. (1980). *Questions of intonation*. London: Routledge.
- Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.
- Buhmann, J., Caspers, J., Heuven, V. v., Hoekstra, H., Martens, J.-P., & Swerts, M. (2002). Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the Spoken Dutch Corpus. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC) 2002* (pp. 779-785), Las Palmas, Spain.
- Büring, D. (1997). The Meaning of Topic and Focus: The 59th Street Bridge Accent. London: Routledge.
- Büring, D. (2003). On D-trees, beans, and B-accent. *Linguistics & Philosophy*, 26(5), 511-545.
- Büring, D. (2007). Intonation, Semantics and Information Structure. In G. Ramchand & C. Reiss (Eds.), *The Oxford Handbook of Linguistic Interfaces* (pp. 445-474). Oxford: Oxford University Press.
- Calhoun, S. (2006). Information Structure and the Prosodic Structure of English: a Probabilistic Relationship. Unpublished PhD Thesis, University of Edinburgh, Edinburgh.
- Carlson, R., Hirschberg, J., & Swerts, M. (2005). Cues to upcoming Swedish prosodic boundaries: Subjective judgment studies and acoustic correlates. *Speech Communication*, 46(3-4), 326-333.
- Caspers, J. (2003). Local speech melody as a limiting factor in the turn-taking system in Dutch. *Journal of Phonetics*, 31, 251-276.
- Chafe, W. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 35-53). Norwood: Ablex.
- 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. N. Li (Ed.), *Subject and Topic* (pp. 25-55). New York: Academic Press.

References

- Chafe, W. L. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and Grounding in Discourse* (pp. 21-51). Amsterdam: John Benjamins.
- Chafe, W. L. (1994). *Discourse, Consciousness and Time*. Chicago/London: The University of Chicago Press.
- Chen, A. (2007). Intonational realisation of topic and focus by Dutch-acquiring 4-to-5-year-olds. In *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)* (pp. 1553-1556), Saarbrücken.
- Chen, A., Den Os, E., & De Ruiter, J. P. (2007). Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech. *Linguistic Review*, 23(2-3), 317-344.
- Cheng, H. Y., Murdoch, B. E., & Goozée, J. V. (2007). Temporal features of articulation from childhood to adolescence: An electropalatographic investigation. *Clinical Linguistics and Phonetics*, 21(6), 481-499.
- Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84, 73-111.
- Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi & B. L. Webber & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). Cambridge: Cambridge University Press.
- Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of verbal learning and verbal behavior*, 22, 245-285.
- Cohen, A., & 't Hart, J. (1967). On the anatomy of intonation. *Lingua*, 19, 177-192.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Physiological Measurements*, 20, 37-46.
- Cole, J., Mo, Y., & Baek, S., (2008). *The role of syntactic structure in guiding prosody perception in spontaneous speech*. Paper presented at the Workshop on Experimental and Theoretical Advances in Prosody, Cornell University.
- Couper-Kuhlen, E., & Selting, M. (1996). Towards an interactional perspective on prosody and a prosodic perspective on interaction. In E. Couper-Kuhlen

References

- & M. Selting (Eds.), *Prosody in interaction* (pp. 11-56). Cambridge: Cambridge University Press.
- Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.
- Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge: Cambridge University Press.
- D'Odorico, L., & Carubbi, S. (2003). Prosodic characteristics of early multi-word utterances in Italian children. *First Language*, 23(1), 97-116.
- Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, 47(2), 292-314.
- Dainora, A., (2002a). *Does intonational meaning come from tones or tunes? Evidence against a compositional approach*. Paper presented at the the 1st Conference on Speech Prosody, Aix-en-Provence, France.
- Dainora, A. (2002b). Modeling intonation in English: A probabilistic approach to phonological competence. In L. Goldstein & D. H. Whalen & C. T. Best (Eds.), *Laboratory Phonology 8*. Berlin: Mouton de Gruyter.
- Dankovicová, J., Pigott, K., Wells, B., & Peppé, S. (2004). Temporal markers of prosodic boundaries in children's speech production. *Journal of the International Phonetic Association*, 34(1), 17-36.
- De Cat, C. (2003). Syntactic manifestations of very early pragmatic competence. In *Proceedings of the 27th annual Boston University Conference on Language Development (BUCLD)* (pp. 209-219), Somerville, MA.
- De Cat, C. (2009). Experimental evidence for preschoolers' mastery of topics. Unpublished manuscript, Leeds.
- De Ruiter, J. P. (2008). Polyfit - Fitting Nth order orthogonal polynomials on weighted data sets (Version 1.0). Nijmegen. Available from jan.deruiter@uni-bielefeld.de.
- De Ruiter, L. E. (2009). The prosodic marking of topical referents in the German "Vorfeld" by children and adults. *The Linguistic Review*, 26(2-3), 329-354.
- Delattre, P. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics*, 4(183-198).

References

- Dombrowski, E. (2003). Semantic features of accent contours: Effects of F0 peak position and F0 time shape. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)* (pp. 1217-1220), Barcelona.
- Drenhaus, H., & Féry, C. (2008). Animacy and child grammar: An OT account. *Lingua*, 118, 222-244.
- Eady, S. J., Cooper, W. E., Klouda, G. V., Mueller, P. R., & Lotts, D. W. (1986). Acoustical characteristics of sentential focus: narrow vs. broad and single vs. dual focus environments. *Language and Speech*, 29(3), 233-251.
- Emslie, H. C., & Stevenson, R. J. (1981). Pre-school children's use of articles in definite and indefinite expressions. *Journal of Child Language*, 8(2), 313-328.
- Erteschik-Shir. (2007). *The syntax/discourse interface: Information Structure* (Vol. 8). Oxford: Oxford University Press.
- Ferrand, C. T. (2002). Harmonics-to-noise ratio: An index of vocal ageing. *Journal of Voice*, 16(4), 480-487.
- Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, 36, 680-703.
- Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*, 79(1), 126-138.
- Fillmore, C. (1982). Frame Semantics. In L. S. o. Korea (Ed.), *Linguistics in the Morning Calm* (pp. 111-138). Hanshin.
- Flavell, J. H. (1974). The development of inferences about others. In T. Mischel (Ed.), *Understanding other persons*. Oxford: Blackwell, Basil & Mott.
- Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: further evidence for the Level 1 - Level 2 distinction. *Developmental Psychology*, 17(1), 99-103.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378-382.
- Frey, W. (2004). A medial topic position for German. *Linguistische Berichte*, 198, 153-190.
- Fujisaki, H. (1992). Modelling the process of fundamental frequency contour generation. In Y. Tokura & E. Vatikiotis-Bateson & Y. Sagisaka (Eds.), *Speech*

References

- perception, production and linguistic structure* (pp. 313-328). Amsterdam: IOS Press.
- Gelman, A., Hill, J., & Yajima, M. (2009, 15-07-2009). *Why we (usually) don't have to worry about multiple comparisons*. Available: <http://arxiv.org/abs/0907.2478v1> [2009, 07-09].
- Gelman, A., & Tuerlinckx, F. (2000). Type S error rates for classical and Bayesian single and multiple comparison procedures. *Computational Statistics, 15*, 373-390.
- Ginsburg, H. (1997). *Entering the child's mind*. Cambridge: Cambridge University Press.
- Givón, T. (1976). Topic, pronoun, and grammatical agreement. In C. Li (Ed.), *Subject and topic* (pp. 151-188). New York: Academic Press.
- Grabe, E. (1998). *Comparative Intonational Phonology: English and German* (Vol. 7). Wageningen: Ponsen en Looien.
- Grabe, E. (2002). *The IViE labelling guide*, [Website]. Available: <http://www.phon.ox.ac.uk/IViE/guide.html> [2009, 06.03.].
- Grabe, E. (2004). Intonational variation in English. In P. Gilles & J. Peters (Eds.), *Regional variation in intonation* (pp. 9-13). Tübingen: Niemeyer.
- Grabe, E., Kochanski, G., & Coleman, J. (2007). Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and Speech, 50*(3), 281-310.
- Grice, M., & Baumann, S. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte, 191*, 267-298.
- Grice, M., Baumann, S., & Benz Müller, R. (2005). German Intonation in Autosegmental-Metrical Phonology. In J. Sun-Ah (Ed.), *Prosodic typology. The Phonology of intonation and phrasing* (pp. 55-83). Oxford: Oxford University Press.
- Grice, M., Baumann, S., & Jagdfeld, N. (2007). Evidence for tonal identity from peak scaling under pitch span variations. In *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)* (pp. 977-980), Saarbrücken.
- Grice, M., Reyelt, M., Benz Müller, R., Mayer, J., & Batliner, A. (1996). Consistency in transcription and labelling of German intonation with

References

- GToBI. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP)* (pp. 1716-1719), Philadelphia.
- Grice, M., Savino, M., & Refice, M. (1997). The intonation of questions in Bari Italian: Do speakers replicate their spontaneous speech when reading? *Phonus*, 3, 1-7.
- Gundel, J. (1999). Topic, focus and the grammar pragmatics interface. In *Proceedings of 23rd Annual Penn Linguistics Colloquium* (pp. 185-200).
- Gundel, J., & Fretheim, T. (2004). Topic and Focus. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 175-196). Malden, MA: Blackwell.
- Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language*, 69, 274-307.
- Gussenhoven, C. (1984). *On the Grammar and Semantics of Sentence Accents*. Dordrecht, Holland; Cinnaminson, N.J., U.S.A.: Foris Publications.
- Gussenhoven, C. (2005). Transcription of Dutch Intonation. In S.-A. Jun (Ed.), *Prosodic typology and transcription: A unified approach* (pp. 118-145). Oxford: Oxford University Press.
- Halliday, M. A. K. (1967a). *Intonation and grammar in British English*. The Hague: Mouton.
- Halliday, M. A. K. (1967b). Notes on Transitivity and Theme in English - Part 2. *Journal of Linguistics*, 3(2), 177-274.
- Halliday, M. A. K. (1983). *Learning how to mean - Explorations in the Development of Language* (Repr. ed.). London: Arnold.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English* (Vol. 9). London: Longman.
- Hawkins, J. (1978). *Definiteness and indefiniteness: a study in reference and grammaticality prediction*. London: Croom Helm.
- Heim, I. (1982). File change semantics and the familiarity theory of definiteness. In R. Bäuerle & C. Schwarye & A. von Stechow (Eds.), *Meaning, use and the interpretation of language* (pp. 164-189). Berlin: De Gruyter.
- Hickmann, M. (1982). *The development of narrative skills: pragmatic and metapragmatic aspects of discourse cohesion*. Unpublished doctoral dissertation, University of Chicago, Chicago.

References

- Hickmann, M. (2003). *Children's Discourse* (Vol. 98). Cambridge: Cambridge University Press.
- Hickmann, M., Hendriks, H. P. J. M., Roland, F., & Liang, J. C. P. (1996). The marking of new information in children's narratives: a comparison of English, French, German and Mandarin Chinese. *Journal of Child Language*, 23(3), 591-619.
- Hinkle, D., Wiersma, W., & Jurs, S. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Boston: Houghton Mifflin Company.
- Hirschberg, J. (1993). Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence*, 63(1/2), 305-340.
- Hirschberg, J., Litman, D., Pierrehumbert, J., & Ward, G. (1987). Intonation and the intentional structure of discourse. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence* (pp. 636-639).
- Hirst, D., Di Cristo, A., & Espreser, R. (1993). Levels of representation and levels of analysis for the description of intonation systems. In M. Horne (Ed.), *Prosody: Theory and Experiment* (pp. 51-88). Dordrecht: Kluwer Academic Press.
- Hockett, C. F. (1958). *A course in modern linguistics*. New York: McMillan.
- Hornby, P. A., & Hass, W. A. (1970). Use of contrastive stress by preschool children. *Journal of Speech and Hearing Research*, 13(2), 395-399.
- Ito, K., & Speer, S. S. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, 58, 541-573.
- Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.
- Jacobs, J. (1982). *Syntax und Semantik der Negation im Deutschen*. München: Fink.
- Jacobs, J. (1997). I-Topikalisierung. *Linguistische Berichte*, 168, 91-133.
- Jacobs, J. (2001). The dimensions of topic-comment. *Linguistics*, 39(4), 641-681.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434-446.

References

- Johnson, K. (2004). Massive reduction in conversational American. In K. Yoneyama & K. Maekawa (Eds.), *Spontaneous speech: Data and analysis* (pp. 29-54). Tokyo: The National Institute for the Japanese Language.
- Jun, S.-A. (Ed.). (2005). *Prosodic typology: the phonology of intonation and phrasing*. Oxford: Oxford University Press.
- Kail, M., & Hickmann, M. (1992). French children's ability to introduce referents in narratives as a function of mutual knowledge. *First Language*, 12, 73-94.
- Karmiloff-Smith, A. (1979). *A functional approach to child language* (Vol. 24). Cambridge: Cambridge University Press.
- Karmiloff-Smith, A. (1981). The grammatical marking of thematic structure in the development of language production. In W. Deutsch (Ed.), *The Child's Construction of Language* (Vol. 3, pp. 121-147). London; New York: Academic Press.
- Katz, W. F., Beach, C. M., Jenouri, K., & Verma, S. (1996). Duration and fundamental frequency correlates of phrase boundaries in productions by childrens and adults. *Journal of the Acoustical Society of America*, 99(5), 3179-3191.
- Keenan, E. O., & Klein, E. (1975). Coherency in Children's Discourse. *Journal of Psycholinguistic Research*, 4(4), 365-378.
- Kendall, M. G., & Babington Smith, B. (1939). The problem of m rankings. *The Annals of Mathematical Statistics*, 10(3), 275-287.
- Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, 3(3), 129-140.
- Klein, W. (1986). *Second language acquisition*. Cambridge: Cambridge University Press.
- Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: fundamental frequency lends little. *Journal of the Acoustical Society of America*, 118(2), 1038-1054.
- Kohler, K. (1991). Terminal intonation patterns in single-accent utterances of German: phonetics, phonology and semantics. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)*, 25, 115-185.

References

- Kreiman, J. (1982). Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics*, 10, 163-175.
- Krifka, M. (2007a). Basic Notions of Information Structure. In C. Féry & G. Fanselow & M. Krifka (Eds.), *The Notions of Information Structure* (Vol. 6, pp. 13-55). Potsdam: Universitätsverlag Potsdam.
- Krifka, M. (2007b). Functional similarities between bimanual coordination and topic/comment structure. In S. Ishihara & S. Jannedy & A. Schwarz (Eds.), *Working Papers of the SFB 632, Interdisciplinary Studies on Information Structure* (Vol. 8, pp. 39-59). Potsdam: Universitätsverlag Potsdam.
- Kruijff, G.-J. (2001). A Categorical-Modal Architecture of Informativity: Dependence Grammar Logic and Information Structure. Unpublished PhD thesis, Charles University, Prague.
- Kruijff-Korbayová, I., & Steedman, M. (2003). Discourse and information structure. *Journal of Logic, Language and Information*. Special Issue on Discourse and Information Structure, 12(3), 249-259.
- Kügler, F. (2007). The intonational Phonology of Swabian and Upper Saxon. (Vol. 515). Tübingen: Niemeyer.
- Kuno, S. (1972). Functional sentence perspective: A case study from Japanese and English. *Linguistic Inquiry*, 3, 269-336.
- Laan, G. P. M. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication*, 22(1), 43-65.
- Ladd, D. R. (1980). *The structure of intonational meaning: evidence from English*. Bloomington: Indiana University Press.
- Ladd, D. R. (1983). Phonological features of intonational peaks. *Language*, 59(4), 721-759.
- Ladd, D. R. (1996). *Intonational Phonology*. Cambridge [England]; New York, NY, USA: Cambridge University Press.
- Ladd, D. R. (2008). *Intonational phonology* (2nd ed. Vol. 119). Cambridge: Cambridge University Press.
- Lambrecht, K. (1994). *Information Structure and Sentence Form*. Cambridge; New York, NY, USA: Cambridge University Press.

References

- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.
- Lehiste, I. (1979). Sentence boundaries and paragraph boundaries - perceptual evidence. In *The Element: a Parasession on Linguistic Units and Levels*, Chicago Linguistic Society, *15*, 99-109.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41-104.
- Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*(1), 339-359.
- Liberman, M., & Pierrehumbert, J. B. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. T. Oehrle (Eds.), *Language Sound Structure: Studies in Phonology Presented to Morris Halle by his Teacher and Students* (pp. 157-233). Boston: MIT Press.
- Lieberman, P. (1967). *Intonation, perception, and language*. Cambridge: MIT Press.
- Local, J. (1980). Modelling intonational variability in children's speech. In S. Romaine (Ed.), *Sociolinguistic variation in speech communities*. London: Edward Arnold.
- Loeb, D. F., & Allen, G. D. (1993). Preschoolers Imitation of Intonation Contours. *Journal of Speech and Hearing Research*, *36*(1), 4-13.
- Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press.
- MacWhinney, B. (2000). *The CHILDES project - tools for analyzing talk*. Mahwah, NJ: Laurence Erlbaum Associates.
- MacWhinney, B., & Bates, E. (1978). Sentential Devices for Conveying Givenness and Newness: A Cross-Cultural Developmental Study. *Journal of Verbal Learning and Verbal Behavior*, *17*, 539-558.
- Maratsos, M. P. (1974). Preschool children's use of definite and indefinite articles. *Child Development*, *45*, 446-455.
- Maratsos, M. P. (1976). *The use of definite and indefinite reference in young children*. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.
- Martin, J. G. (1970). On judging pauses in spontaneous speech. *Journal of Verbal Learning and Verbal Behavior*, *9*(1), 75-78.

References

- Mehlhorn, G. (2001). Produktion und Perzeption von Hutkonturen im Deutschen. *Linguistische Arbeitsberichte*, 77, 31-57.
- Menard, S. W. (1995). *Applied Logistic Regression Analysis* (Vol. 106). Thousand Oaks, CA: Sage Publications.
- Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, 32, 543-563.
- Mo, Y. (2008). Duration and intensity as perceptual cues for naïve listeners' prominence and boundary perception. In *Proceedings of the 4th Conference on Speech Prosody* (pp. 739-742), Campinas, Brazil.
- Mo, Y., Cole, J., & Lee, E.-K. (2008). Naïve listeners' prominence and boundary perception. In *Proceedings of the 4th Conference on Speech Prosody* (pp. 735-738), Campinas, Brazil.
- Möbius, B., & Jilka, M. (2007). Effects of syllable structure and nuclear pitch accents on peak alignment: A corpus-based analysis. In *Proceedings of 16th International Congress of the Phonetic Sciences (ICPhS) 2007* (pp. 1173-1176), Saarbrücken.
- Moll, H., Koring, C., Carpenter, M., & Tomasello, M. (2006). Infants Determine Others' Focus of Attention by Pragmatics and Exclusion. *Journal of Cognition and Development*, 7(3), 411-430.
- Moll, H., & Tomasello, M. (2006). Level I perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, 24, 603-613.
- Moll, H., & Tomasello, M. (2007). How 14- and 18-Month-Olds Know What Others Have Experienced. *Developmental Psychology*, 43(2), 309-331.
- Monschau, J. (2004). Input and output in oral reading in English: the interaction of syntax, semantic-pragmatics and intonation. Unpublished PhD thesis, Rheinische Friedrich-Wilhelms-Universität, Bonn.
- Müller, A., Höhle, B., Schmitz, M., & Weissenborn, J. (2006). Focus-to-stress alignment in 4 to 5-year-old German-learning children. In *Proceedings of GALA* (pp. 393-407), Cambridge.
- Nolan, F. (2003). Intonational equivalence: An experimental evaluation of pitch scales. In *Proceedings of XVth International Congress of Phonetic Sciences* (pp. 771-774), Barcelona.

References

- Ohala, J., & Ewan, W. G. (1973). Speech of pitch change. *Journal of the Acoustical Society of America*, 53, 345 (A).
- Oller, D. K. (1973). The effect of position in utterance on speech segment duration. *Journal of the Acoustical Society of America*, 54, 1235-1247.
- Ota, M. (2008). Children's production of word accents in Swedish revisited. *Phonetica*, 63, 320-246.
- Patel, R., & Grigos, M. I. (2006). Acoustic characterization of the question-statement contrast in 4, 7 and 11 year-old children. *Speech Communication*, 48, 1308-1318.
- Paul, H. (1880). *Prinzipien der Sprachgeschichte*. Leipzig.
- Pierrehumbert, J. B. (1980). *The Phonetics and Phonology of English intonation*. Unpublished PhD Thesis, MIT, Bloomington.
- Pierrehumbert, J. B., & Hirschberg, J. (1990). The Meaning of Intonational Contours in the Interpretation of Discourse. In P. R. Cohen & J. Morgan & M. E. Pollack (Eds.), *Intentions in Communication* (pp. 271-311). Cambridge: MIT Press.
- Pijper, J. R. d., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America*, 96(4), 2037-2047.
- Pitrelli, J. F., Beckman, M., & Hirschberg, J. (1994). Evaluation of prosodic transcription labelling reliability in the ToBI framework. In *Proceedings of the 3rd International Conference on Spoken Language Processing (ICSLP)* (pp. 123-126), Yokohama.
- Potamianos, A., & Narayanan, S. (1998, May 1998). Spoken dialog systems for children. In *Proceedings of Internat. Conf on Acoust., Speech, and Signal Process.* (pp. 197-201), Seattle, Washington.
- Pratt, M. W., & MacKenzie-Keating, S. (1985). Organizing stories: effects of development and task difficulty on referential cohesion in narrative. *Developmental Psychology*, 21(2), 350-356.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1988). *Numerical Recipes in C: the Art of Scientific Computing*. Cambridge: Cambridge University Press.

References

- Price, P., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, *90*, 2956-2970.
- Prieto, P., D'Imperio, M., & Fivela Gili, B. (2005). Pitch accent alignment in Romance: primary and secondary associations with metrical structure. *Language and Speech*, *48*(4), 359-396.
- Prieto, P., Santen, J. v., & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, *23*, 429-451.
- Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics* (pp. 223-256). New York: Academic Press.
- R Development Core Team. (2008). R: A language and environment for statistical computing (Version 2.6.2). Vienna: R Foundation for Statistical Computing.
- Randolph, J. J., (2005). *Free-marginal multirater kappa: an alternative to Fleiss' fixed-marginal multirater kappa*. Paper presented at the Joensuu Learning and Instruction Symposium, Joensuu, Finland.
- Rathcke, T., & Harrington, J. (in press). The variability of early accent peaks in Standard German, *Papers in Laboratory Phonology* (Vol. 10). Cambridge: Cambridge University Press.
- Reinhart, T. (1981). Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, *27*, 53-94.
- Sanderman, A. A. (1996). Prosodic phrasing. Eindhoven.
- Schegloff, E. A. (1996). Turn organization: One intersection of grammar and interaction. In E. Ochs & E. A. Schegloff & S. A. Thompson (Eds.), *Interaction and Grammar* (pp. 52-133). Cambridge: Cambridge University Press.
- Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361-382.
- Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J. M., & Stahl, S. A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology*, *96*(119-129).

References

- Scobbie, J. M., Gibbon, F., Hardcastle, W. J., & Fletcher, P. (1996). Covert contrast as a stage in the acquisition of phonetics and phonology. *QMC Working papers in Speech and Language Sciences, 1*, 43-62.
- Scollon, R. (1979). A real early stage: an unzipped condensation of a dissertation on child language. In E. Ochs & B. Schieffelin (Eds.), *Developmental pragmatics*. New York: Academic Press.
- Scott, W. A. (1955). Reliability of content analysis: The case of nominal scale coding. *The Public Opinion Quarterly, 19*(321-325).
- Shattuck-Hufnagel, S., Ostendorf, M., & Ross, K. (1994). Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics, 22*(4), 357-388.
- Shattuck-Hufnagel, S., & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193-247.
- Silverman, K. E., Beckman, M., Pitrelli, J. F., Ostendorf, M., Wrightman, C., Price, P., Pierrehumbert, J. B., & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. In *Proceedings of the 2nd International Conference on Spoken Language Processing (ICSLP)* (pp. 867-870), Banff, Canada.
- Silverman, K. E., & Pierrehumbert, J. B. (1990). The timing of prenuclear high accents in English. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech* (pp. 72-106). Cambridge: Cambridge University Press.
- Sluijter, A. M. C., & Van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America, 100*(4), 2471-2485.
- Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48*, 103-130.
- Snow, D. (1998). Children's imitations of intonation contours: Are rising tones more difficult than falling tones? *Journal of Speech Language and Hearing Research, 41*(3), 576-587.

References

- Snow, D. (2001). Imitation of intonation contours by children with normal and disordered language development. *Clinical Linguistics and Phonetics*, 15(7), 567-584.
- Snow, D., & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua*, 112(12), 1025-1058.
- Sosa, J. M. (1999). La entonación del español: Su estructura fónica, variabilidad y dialectología. Madrid: Cátedra.
- Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Pragmatics* (Vol. 9, pp. 315-332). New York: Academic Press.
- Stathopoulos, E. T., & Sapienza, C. M. (1997). Developmental changes in laryngeal and respiratory function with variations in sound pressure level. *Journal of Speech Language and Hearing Research*, 40(3), 595-614.
- Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, 31(4), 649-689.
- Steedman, M. (2007). Information-structural semantics of English intonation. In C. Lee & M. Gordon & D. Büring (Eds.), *Topic and Focus - Cross-Linguistic Perspectives on Meaning and Intonation* (pp. 245-264). Dordrecht: Springer.
- Streefkerk, B. M., Pols, L. C. W., & Bosch, L. F. M. t. (1997). Prominence in read aloud sentences, as marked by listeners and classified automatically. In *Proceedings of IFA* (pp. 101-116), Amsterdam.
- Streefkerk, B. M., Pols, L. C. W., & Bosch, L. F. M. t. (1998). Automatic classification of prominence (as defined by listeners' judgements) in read aloud Dutch sentences. In *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP)* (pp. 683-686), Sydney.
- Sturm, J. A., & Seery, C. H. (2007). Speech and articulatory rates of school-age children in conversation and narrative contexts. *Language Speech and Hearing Services in Schools*, 38(1), 47-59.
- Swerts, M., & Heldner, M. (1996). F0 declination in spontaneous and read-aloud speech. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP)* (pp. 1501-1504), Philadelphia, PA.

References

- Syrdal, A. K., & McGory, J. (2000). Inter-transcriber reliability of ToBI prosodic labeling. In *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP)* (pp. 235-328), Beijing, China.
- Taylor, P. (2000). Analysis and synthesis of intonation using the Tilt model. *Journal of the Acoustical Society of America*, 107(3), 1697-1714.
- Terken, J. M. B. (1984). The distribution of pitch accents in instructions as a function of discourse structure. *Language and Speech*, 27(3), 269-289.
- Terken, J. M. B., & Hirschberg, J. (1994). Deaccentuation of Words Representing 'Given' Information: Effects of Persistence of Grammatical Role and Surface Position. *Language and Speech*, 37, 125-145.
- Terken, J. M. B., & Nootboom, S. G. (1987). Opposite Effects of Accentuation and Deaccentuation on Verification Latencies for Given and New Information. *Language and Cognitive Processes*, 2, 145-163.
- Tomasello, M., & Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what is new for other persons. *Developmental Psychology*, 39(5), 906-912.
- Truckenbrodt, H. (2007). Upstep on edge tones and on nuclear accents. In C. Gussenhoven & T. Riad (Eds.), *Tones and tunes: experimental studies in word and sentence prosody*. (pp. 349-386). Berlin: Mouton.
- Uhmann, S. (1991). Fokusphonologie. Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie. Tübingen: Niemeyer.
- Vallduví, E. (1992). *The Informational Component*. New York: Garland.
- Vallduví, E., & Engdahl, E. (1996). The Linguistic Realisation of Information Packaging. *Linguistics*, 34, 459-519.
- van den Berg, R., Gussenhoven, C., & Rietveld, A. (1992). Downstep in Dutch: implications for a model. In G. R. Docherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology 2: Gesture, Segment, Prosody* (Vol. 2, pp. 335-359). Cambridge.
- von Heusinger, K. (1999). *Intonation and Information Structure*. Unpublished State doctorate thesis (Habilitationsschrift), University of Konstanz, Konstanz.

References

- von Stutterheim, C., & Carroll, M. (2005). Subjektwahl und Topikkontinuität im Deutschen und Englischen. *Zeitschrift für Literaturwissenschaft und Linguistik*, 139, 7-27.
- Warden, D. A. (1976). The influence of context on children's use of identifying expressions and references. *British Journal of Psychology*, 67(1), 101-112.
- Watson, D., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science*, 32(7), 1232-1244.
- Weil, H. (1844). *De l'Ordre des Mots Dans les Langues Anciennes Comparées aux Langues Modernes* (Thèse française). Paris: Joubert.
- Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language*, 31(4), 749-778.
- Wieman, L. A. (1976). Stress patterns in early child language. *Journal of Child Language*, 3(2), 283-286.
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91, 1707-1717.
- Wittek, A., & Tomasello, M. (2005). Young children's sensitivity to listener knowledge and perceptual context in choosing referring expressions. *Applied Psycholinguistics*, 26, 541-558.
- Wonnacott, E., & Watson, D. G. (2008). Acoustic emphasis in 4-year-olds. *Cognition*, 107(3), 1093-1101.
- Wunderlich, D. (1988). Der Ton macht die Melodie - Zur Phonologie der Intonation des Deutschen. In H. Altmann (Ed.), *Intonationsforschungen* (Vol. 200, pp. 1-40). Tübingen: Niemeyer.
- Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 46, 220-251.
- Xu, Y., & Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America*, 111(3), 1399-1413.
- Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33, 159-197.
- Yule, G. (1981). New, current and displaced entity reference. *Lingua*, 55, 41-52.

Appendices

A Picture stories

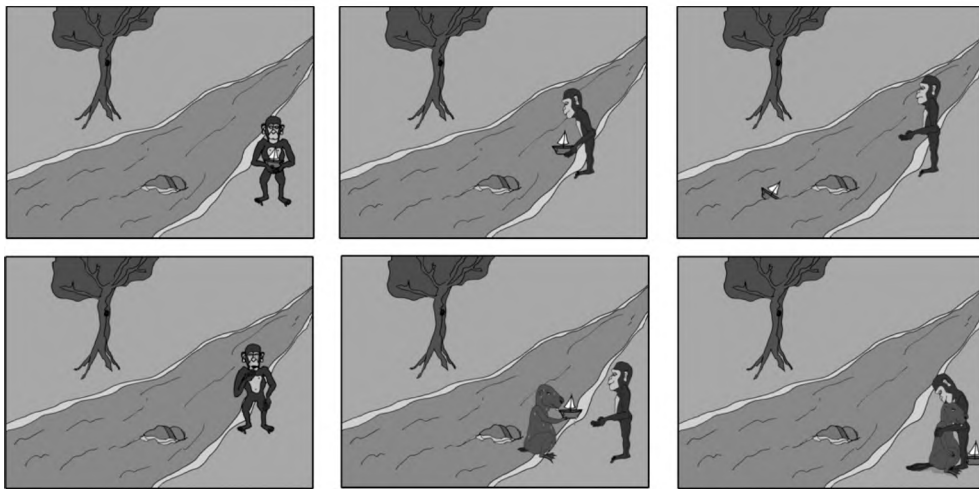


Figure 27: Picture story for the elicitation of the target referent *Biber* in the given condition (picture 6).

Appendix A Picture stories

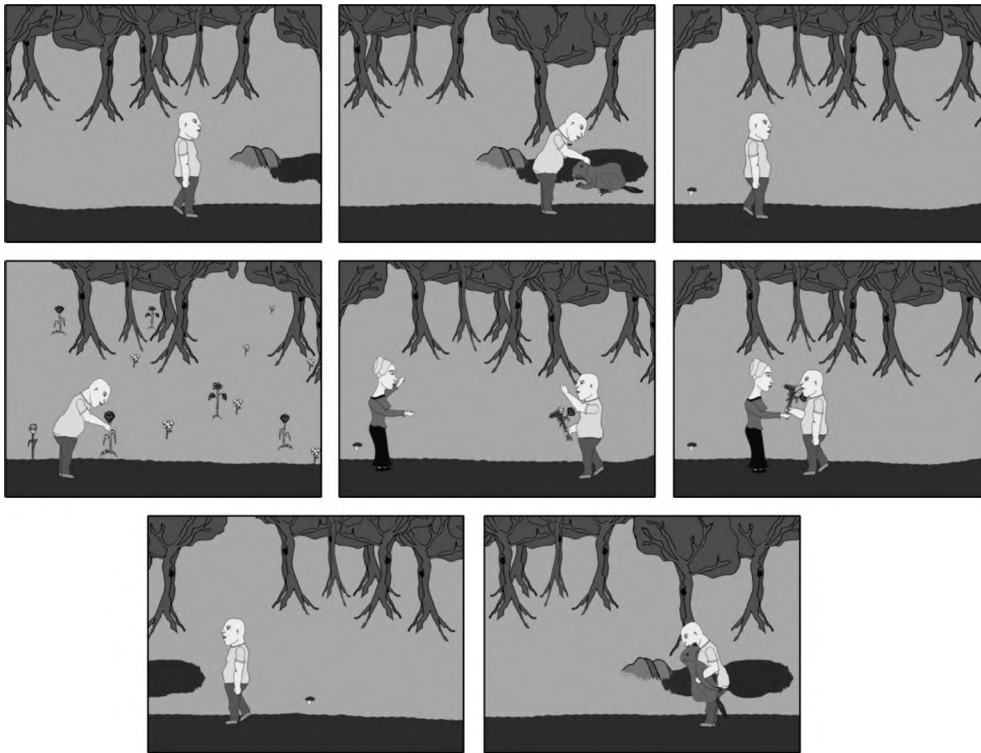


Figure 28: Picture story for the elicitation of the target referent *Biber* in the new condition (picture 2) and in the accessible condition (picture 8).

Appendix A Picture stories

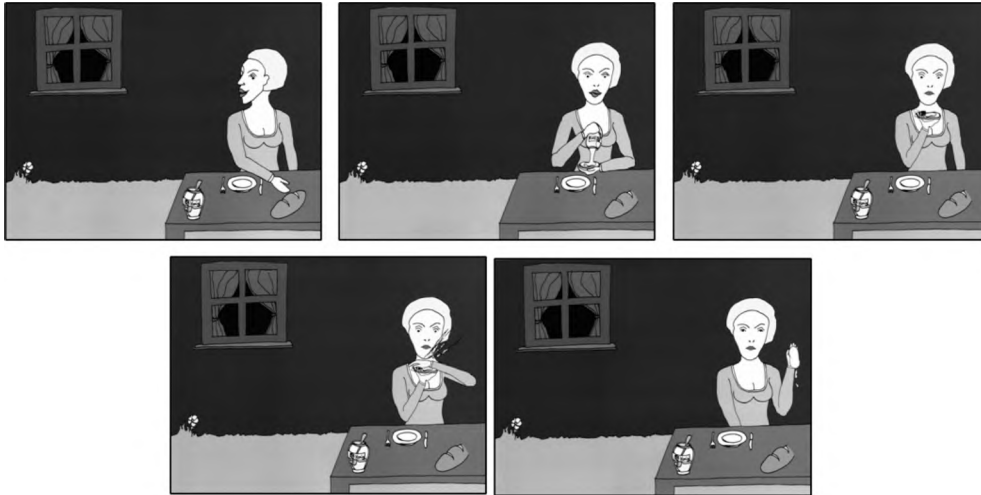


Figure 29: Picture story for the elicitation of the target referent *Biene* in the given condition (picture 4).

Appendix A Picture stories

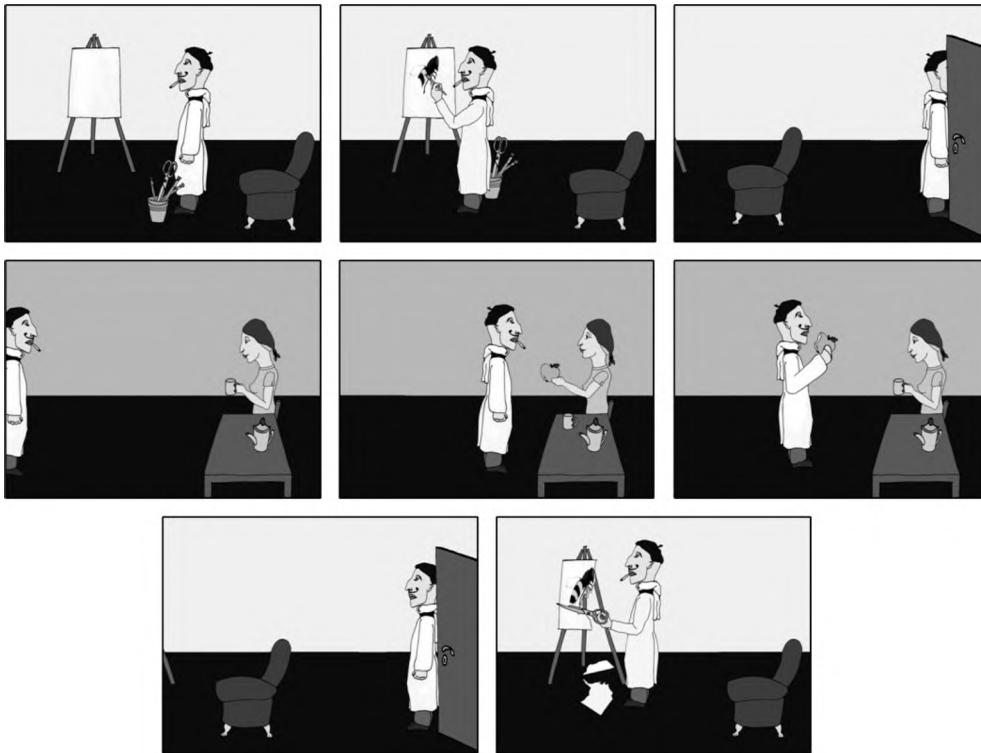


Figure 30: Picture story for the elicitation of the target referent *Biene* in the new condition (picture 2) and in the accessible condition (picture 8).

Appendix A Picture stories

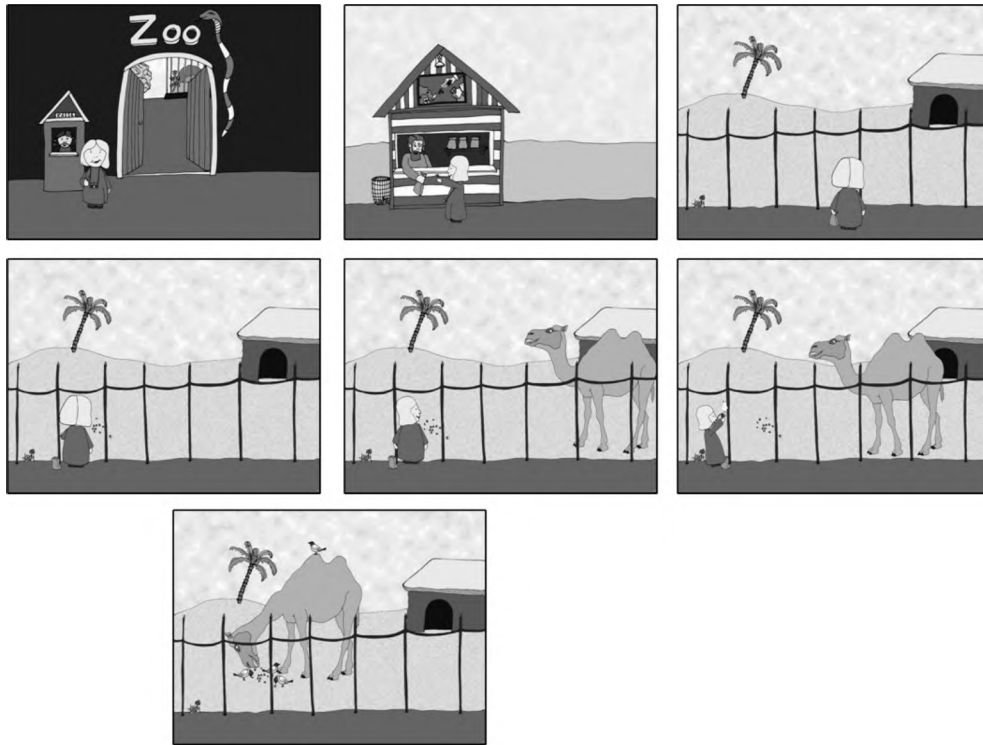


Figure 31: Picture story for the elicitation of the target referent *Kamel* in the given condition (picture 6).

Appendix A Picture stories

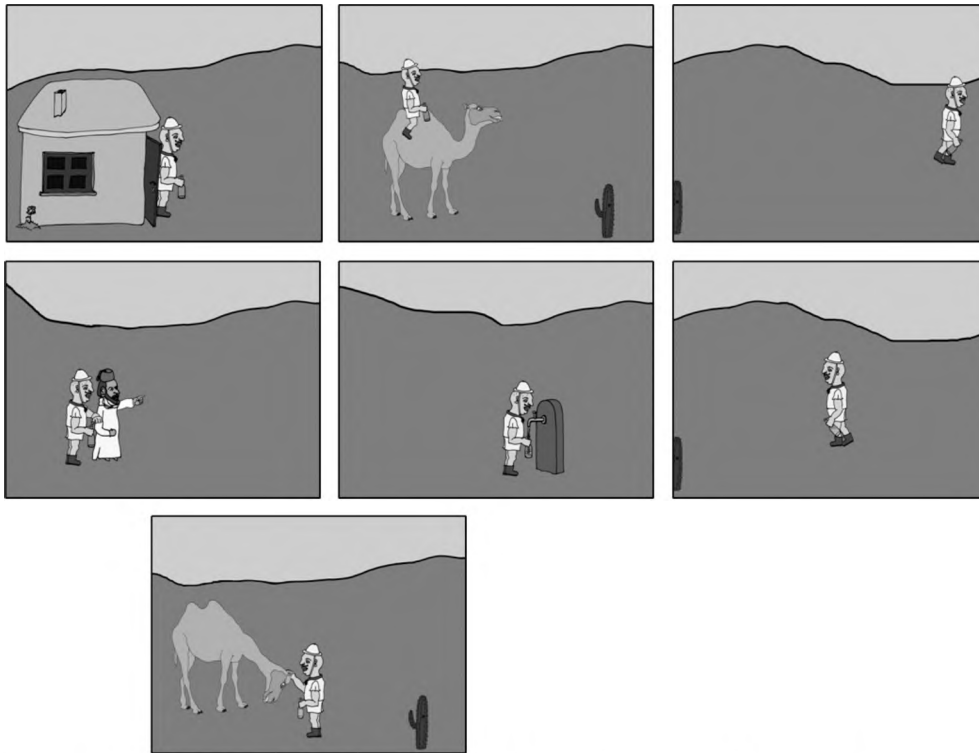


Figure 32: Picture story for the elicitation of the target referent *Kamel* in the new condition (picture 2) and in the accessible condition (picture 8).

Appendix A Picture stories

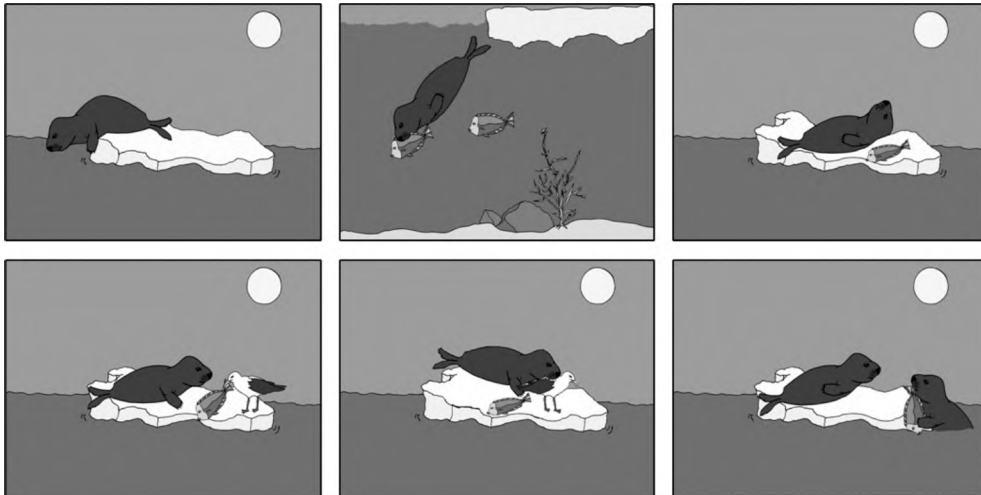


Figure 33: Picture story for the elicitation of the target referent *Möwe* in the given condition (picture 5).

Appendix A Picture stories

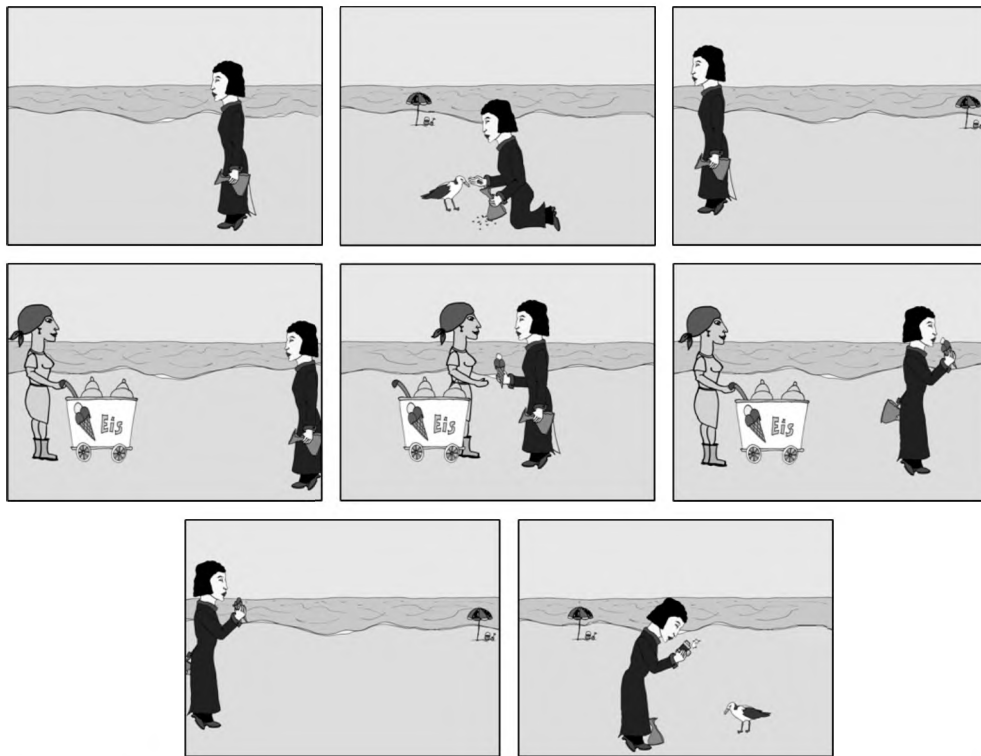


Figure 34: Picture story for the elicitation of the target referent *Möwe* in the new condition (picture 2) and in the accessible condition (picture 8).

Appendix A Picture stories



Figure 35: Picture story for the elicitation of the target referent *der Sandmann* in the new condition (picture 2).



Figure 36: Picture story for the elicitation of the target referent *der Weihnachtsmann* in the new condition (picture 2).

B Pitch accent label consistency check

This section first describes the data set used for checking the consistency of the pitch accent labels assigned to the data in chapter 4. Then the defining features for each accent type are listed. From this, a number of predictions for differences between the accent types are derived. After an explanation of the statistical method, the results are presented and discussed.

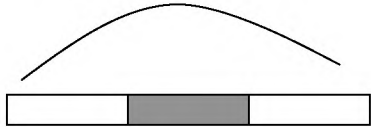
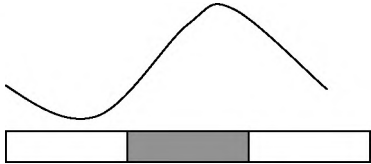
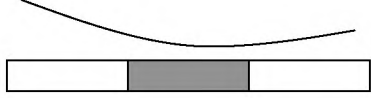
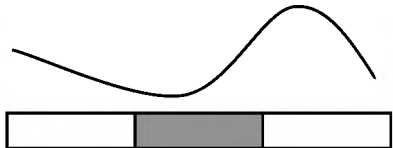
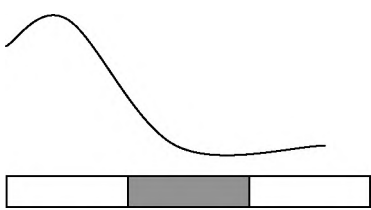
1 Material

In most cases, the non-target words contained fewer sonorant segments than the target words (e.g., Eichhörnchen, [ˈʔaɪçhœʁnçən] ‘squirrel’ instead of *Biber* ‘beaver’, or *Hornisse* [hɔɐˈnɪsə] ‘hornet’ instead of *Biene* ‘bee’). This makes pitch tracking, and consequently phonetic analysis much more difficult. Because they are phonetically maximally comparable, the phonetic analysis was therefore performed on the target words. From these, 293 utterances were selected in which the target word (*Biene, Biber, Möwe, Kamel*) had been produced in phrase-final position. The decision of whether an item was phrase-final or not was based on the results from the perception study described in chapter 7. Of the 293 utterances, 80 came from the five-year-old group, 113 from the seven-year-olds and 100 from the adult group.

2 Types of pitch accents and boundary tones occurring in the corpus

Using the GToBI annotation scheme, seven different pitch accent types the speakers used could be identified. In addition, a substantial proportion of referents were deaccented. Schematic representations of the seven pitch accent types are provided in Table 25 below, along with a description of their respective characteristic features as stated in Grice et al. (2005).

Appendix B Pitch accent label consistency check

Symbol	Schematic representation	Characteristic features
H*		<ul style="list-style-type: none"> - 'peak accent' - perceived as relatively high - may be preceded by shallow rise -
L+H*		<ul style="list-style-type: none"> - 'rise from low up to peak accent' - peak preceded by low pitch target, sharp rise in accented syllable - peak often late in accented syllable
L*		<ul style="list-style-type: none"> - 'low accent' - accented syllable is local pitch minimum - low in the speaker's range
L*+H		<ul style="list-style-type: none"> - 'valley accent plus rise' - low target within accented syllable followed by a rise, starting late in accented syllable - peak on next syllable (or later)
H+L*		<ul style="list-style-type: none"> - 'step-down from high to low accent' - accented syllable low at or very near the bottom of the speaker's range - preceded by high pitch target on preceding syllable

Appendix B Pitch accent label consistency check

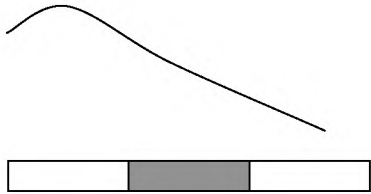
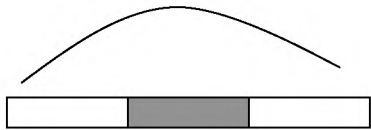
<p>H+!H*</p>		<ul style="list-style-type: none"> - 'step-down from high to mid accent' - accented syllable preceded by higher pitch on preceding syllable - accented syllable around the middle of the range - often continuous fall from the pre-accented syllable through the accented syllable to final syllable in phrase
<p>!H*</p>		<ul style="list-style-type: none"> - a H* accent that is shifted downwards in comparison to a previous H tone in the same intermediate phrase (ip)

Table 25: Schematic representation of the seven GToBI pitch accent types present in the corpus. The grey-shaded area indicates the lexically stressed syllable, the white areas preceding and following it represent the pre- and post-accentual syllable, respectively.

From the characteristic features, predictions can be derived with respect to the way in which one accent should differ from another. For instance, H* and L+H* both feature a high tonal target (a pitch peak). However, a characteristic feature of L+H* accents is a sharp rise in the accentual syllable. We should therefore expect that the slope of the pitch in L+H* accents is steeper than in H* accents. In a similar vein, predictions can be made for other pairs of pitch accents.

Making these predictions is complicated by the fact that pitch accents can be combined with different types of boundary tones. The speakers in the present corpus produced five types (descriptions taken from Grice et al. 2005):

Appendix B Pitch accent label consistency check

- L-%: 'F0 minimum low in the speaker's range'; 'may be followed by drop to extra low'
- L-H%: 'low followed by rise to mid'
- H-%: 'roughly the same F0 value as the peak corresponding to the most recent H tone', high 'plateau'
- !H-%: 'downstepped in relation to previous accentual H tone'
- H-^H%: 'plateau followed by sharp rise at the end of the phrase'

The illustrations in Table 25 describe situations in which there is a downward pitch movement following the pitch accent, as is the case when a pitch accent is followed by the boundary tones L-%, L-H% or !H-%. Identifying high and low targets and calculating parameters like slope is straightforward. The situation is different with high boundary tones (H-% and H-^H%). Since all utterances are phrase-final, the high target of the pitch accent and the high target of the boundary tone will often co-occur on the post-accentual syllable (if there is one), and will in most cases be merged into one pitch maximum at the edge of the phrase. The position and value of this pitch maximum are not informative, because they are aggregates of pitch accent and boundary tone. As a consequence, the pitch excursion and speed of pitch change in H* and L+H* accents, for example, may not differ substantially from each other, as illustrated in Figure 37.

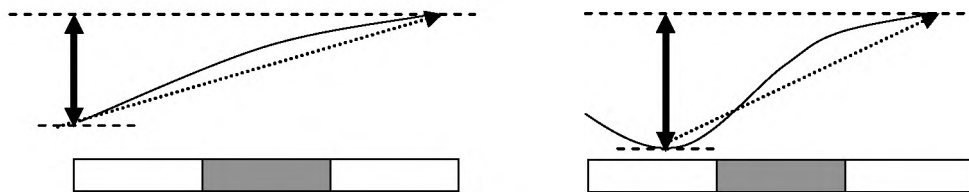


Figure 37: A H* and a L+H* accent followed by a high boundary tone, with the pitch maximum measured at the end of the phrase. The thick black arrows indicate the pitch excursion, the dotted arrow the slope.

Appendix B Pitch accent label consistency check

However, the accent pairs should exhibit differences on the lexically stressed syllable itself and/or the preceding syllable. To stick with the example just given, we should expect the sharp rise in the accented syllable of the L+H* accents to be reflected in a steeper slope when the pitch maximum is measured not at the end of the phrase, but rather at the end of the accented syllable (Figure 38).

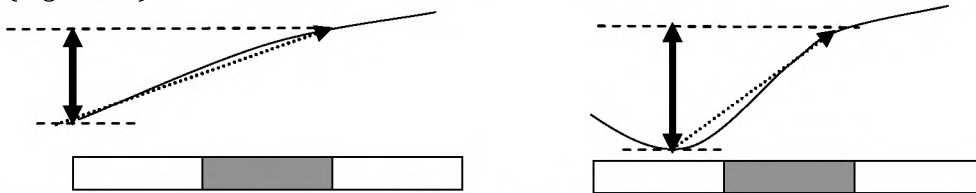


Figure 38: A H* and a L+H* accent followed by a high boundary tone, with pitch maximum measured at the end of the accented syllable. The thick black arrows indicate the pitch excursion, the dotted arrow the slope.

In the case of high boundary tones, predictions are therefore made for the domain up to, but not including, the post-accentual syllable.

3 Predictions

Table 26 contrasts accent pairs, especially those that have been found to be prone to confusion in the Grice et al. (1996) study, and lists the predictions that are made with respect to the phonetic differences we should expect to find between these pairs. Differential predictions are made for accents occurring before low and high boundary tones, respectively.

Appendix B Pitch accent label consistency check

accent pair	low boundary (L-%, L-H%, !H-%)	high boundary (H-%, H- [^] H%)
H* vs. L+H*	higher excursion in L+H* steeper slope in L+H*	higher excursion in L+H* steeper slope in L+H*
L+H* vs. L*+H	pitch minimum aligned later in L*+H pitch maximum aligned later in L*+H	pitch minimum aligned later in L*+H
L* vs. L*+H	--	higher excursion in L*+H steeper slope in L*+H pitch minimum aligned later in L*
H+L* vs. H+!H*	higher excursion in H+L* steeper slope in H+L*	--
H* vs. !H*	Larger pitch difference between maximum and preceding peak in !H*	--
deaccented vs. T* ⁷⁵	higher excursion in T* steeper slope in T*	higher excursion in T* ? steeper slope in T* ?

Table 26: Predicted phonetic differences between several pitch accent types occurring before low boundary tones (L-%, !H-%) and high boundary tones (H-%, H-[^]H%). Where no prediction is made (indicated by a "--" in the cell), this is due to the fact that at least one of the accents of a given pair did not occur in combination with that boundary tone.

4 Annotation and phonetic measurements

The intonational phrase (IP) containing the target word was first segmented at the level of the syllable using visual information from a wide-band spectrogram. The boundary between a nasal and a neighboring vowel (in utterances such as *eine Möwe* [aɪnəmø:və], *Kamel* [kame:l] or *Biene* [bi:nə]) was identi-

⁷⁵ "T*" is used as a generic symbol for any accent type (as apposed to deaccented).

Appendix B Pitch accent label consistency check

fied as the point in the spectrogram where there was a sudden change in spectral distribution of the energy. In vowel-plosive sequences (e.g., *die Biene* [di**b**i:nə]), the beginning of the consonant was determined as the cessation of voicing of the preceding vowel. The beginning of a vowel following a plosive was set at the onset of the aperiodic noise following the release of the closure (release burst).

The position and value of local fundamental frequency (F0) maxima (max) and minima (min) were determined manually for the accents H*, !H*, L+H*, L*, and L*+H. In accents that were followed by a low boundary tone (L-%) or a downstepped boundary tone (!-H%), the domain in which these landmarks were set consisted of the lexically stressed syllable, the preceding syllable and the syllable following it. In accents that were followed by a high boundary tone (H-%) or an upstepped boundary tone (H-^H%), min and max were determined in the pre-accented and the accented syllable only. For H+!H* and H+L*, the high on the pre-accentual syllable was measured. In H+L* accents, the minimum in the accented syllable was taken. In H+!H* accents, a second max on the accented syllable was measured if there was a discernible peak or turning point in that syllable. Otherwise, pitch was measured in the middle of the accented syllable. For H* and !H*, an additional F0 measurement was taken on the preceding high pitch accent. In deaccented target words, min and max were measured wherever they occurred in the three-syllable domain. Figure 39 indicates the measurement points for the various pitch accent type + boundary tone combinations.

Appendix B Pitch accent label consistency check

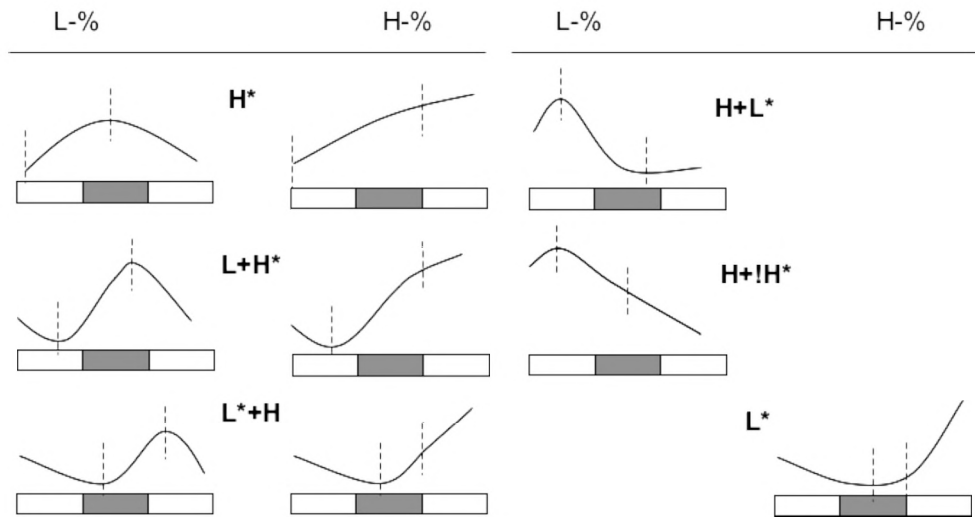


Figure 39: F0 minima and maxima measurements for the various pitch accent types, in combination with high and low boundary tones (where applicable). The grey-shaded area indicates the lexically stressed syllable.

From these F0 landmarks, several derived measurements were made:

- the excursion (F0 max-min; in semitones (st))
- the duration of the rise/fall (in seconds (sec))
- the slope of the rise/fall (excursion/duration of rise/fall, in st/sec)
- the position of the maximum (in % into the accented syllable)
- the position of the minimum (in % into the accented syllable)

5 Statistical procedure

The predictions listed in Table 26 were tested using multilevel regression modeling in R. This method allows incorporating fixed factors with an explanatory value (such as pitch accent type) together with crossed random factors such as subject or item. These linear mixed-effect models (LME) have a number of advantages over traditional methods like ANOVA, such as robustness with respect to missing data (Baayen 2008, p. 290). For each accent pair, I tested whether pitch accent type (henceforth ACCENT) is a significant pre-

Appendix B Pitch accent label consistency check

dictor for the phonetic dimension in question (i.e., pitch excursion, position of F0 maximum etc.). Thus the models contained the parameter in question as the dependent variable and ACCENT as fixed factor. In addition, subject (SUBJECT) and word (WORD) were included as random factors, but excluded if the variance explained by them was effectively zero. If the explained variance was not zero, it was tested whether the random effect parameters are justified in the model. To this end, likelihood ratio tests were applied (Baayen, 2008, p. 275f.). These tests compare the log likelihood (a measure of goodness of fit) of two models with different factor structure (e.g., one with two random factors and one with only one random factor). If the goodness of fit of the model with more factors is not significantly better than that of the model with fewer, the more parsimonious model was chosen. All reported p-values were obtained by estimating the posterior probability of a Markov Chain Monte Carlo (MCMC) simulation with 10,000 runs.

6 Results

The distribution of the various accentuation types (including deaccentuation) in combination with high and low boundary tones in the sample is displayed in Table 27. It is striking that some pitch accent types occur preferably or even exclusively with one type of boundary tone and not the other (Cramer's $V = .645$, $p < .001$). This result is in line with previous observations that the type of pitch accent is a strong predictor for the type of boundary tone, and vice versa (e.g., Dainora, 2002a).

Appendix B Pitch accent label consistency check

		boundary	
		high	low
		Count	Count
Accent	deacc	16	29
	H*	19	40
	L+H*	23	28
	L*	28	0
	L*+H	48	1
	H+L*	0	10
	H+!H*	1	34
	!H*	0	16
	Total	135	158

Table 27: Distribution of accentuation types by boundary tone.

Deaccentuation vs. T*

Before looking at the differences between the various accent types, an analysis was conducted to check whether items that had been labeled “deaccented” differed from items that had been marked as bearing an accent. When a constituent is deaccented there should be only little pitch movement in the three-syllable domain, at least in the case of utterances that end in a low boundary tone. One would expect the pitch to stay approximately level after the last pitch accent of the phrase. Consequently, pitch excursion (henceforth EXC) should be small and/or the slope of the pitch (henceforth SLOPE) should be rather shallow compared to accented items. In the case of a high and especially an upstepped boundary tone, the pitch often rises again on the last syllable, so that EXC and SLOPE should be larger and steeper than were the phrase ends in a low boundary tone. In order to test phonetic differences be-

Appendix B Pitch accent label consistency check

tween accented and deaccented items, EXC and SLOPE in deaccented items were compared to those in all different types of accents.

When a phrase ended with a low boundary tone, deaccented items had a significantly smaller EXC than items labeled H+L* and L+H*. No significant differences in EXC were found for items labeled H*, !H* and H+!H*. The results of the analysis are shown Table 28.⁷⁶

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	2.00	1.38	2.61	0.0001
H*	-0.08	-0.92	0.71	0.8124
L+H*	2.72	1.90	3.63	0.0001
H+L*	5.14	3.93	6.54	0.0001
H+!H*	0.12	-0.74	0.94	0.7882
!H*	-0.84	-1.90	0.16	0.1010

Table 28: Excursion estimates (in st), lower and upper bounds and probability levels for all accentuation types followed by a low boundary tone. As there was only one token, L*+H was excluded. N = 157

When deaccentuation in low-ending phrases is compared to accents in terms of SLOPE, all accents have a significantly steeper slope (see Table 29).

⁷⁶ The second line of the table gives the values for the reference level in the model, the intercept (here: deaccented items). In order to obtain the values for the other levels (e.g., H*), their estimate needs to be added to the estimate of the intercept. The same procedure applies to the upper and lower bounds.

Appendix B Pitch accent label consistency check

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	-2.82	-5.93	0.36	0.0746
H*	11.68	7.47	15.64	0.0001
L+H*	24.40	20.56	29.37	0.0001
H+L*	-29.04	-35.02	-22.45	0.0001
H+!H*	-6.91	-10.88	-2.33	0.0024
!H*	11.34	6.31	16.80	0.0002

Table 29: Slope estimates (in st/sec), lower and upper bounds and probability levels for all accentuation types followed by a low boundary tone. N = 157.

The results are not as clear-cut in the case of high-ending utterances. The accents L*, L*+H and L+H* show a significantly higher EXC than deaccented items. However, this does not hold for H* (see Table 30).

Appendix B Pitch accent label consistency check

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	1.97	-2.33	6.21	0.2788
H*	-0.02	-2.55	2.51	0.9780
L+H*	5.53	3.02	7.87	0.0001
L*	3.20	0.76	5.45	0.0086
L*+H	4.75	2.66	6.90	0.0001

Table 30: Excursion estimates, lower and upper bounds and probability levels for all accentuation types followed by a high boundary tone. Since there was only one token of H+!H*, it was not included in the analysis. N = 134.

The same picture emerges for the SLOPE: While SLOPE is significantly steeper in L*, L*+H and L+H*, H* does not differ significantly from deaccented items in this dimension (see Table 31).

Appendix B Pitch accent label consistency check

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	1.87	-7.84	12.27	0.5778
H*	7.60	-1.95	17.02	0.1162
L+H*	23.22	14.01	32.26	0.0001
L*	19.81	10.15	28.26	0.0001
L*+H	25.70	17.50	33.58	0.0001

Table 31: Slope estimates, lower and upper bounds and probability levels for all accentuation types followed by a high boundary tone. Since there was only one token of H+!H*, it was not included in the analysis. N = 134.

The failure to find a significant difference between deaccented items and H* in both parameters is surprising. A potential explanation for this outcome is the way excursion and slope were measured in deaccented items. Recall that in deaccented items, the lowest and the highest point anywhere within the three-syllable domain were taken as extrema. Given that the pitch would be expected to rise toward the high boundary tone – albeit slightly – even in deaccented items, and given that a deaccented word could also be followed by an upstepped boundary (i.e., a pitch increase on the last syllable), the overall pitch difference is likely to be different from zero. In contrast to this, the pitch maximum in H* accents was taken to be the highest pitch point in the accented syllable, disregarding any pitch movement following it. Even though there is pitch movement on the first two syllables of the measuring domain, this pitch movement need not necessarily be greater than that found in an deaccented word followed by a high or upstepped boundary. In addition, it is possible that the difference between deaccentuation and H* before high boundary tones may be signaled by other acoustic parameters such as intensity and duration,

Appendix B Pitch accent label consistency check

which have been found to be correlates of prominence (e.g., Kochanski et al., 2005). In order to check whether they may be errors in the annotations of H* and deaccented items (before high boundary tones), I went back to the data. The critical items were inspected again and the labels checked together with a second phonetically trained annotator. Both annotators came to the conclusion that the labels were assigned correctly, but agreed that in some items the perception of prominence may be cued by other parameters than F0.

Accent pair contrasts

H* vs. L+H*

In low-ending phrases, L+H* accents had an EXC that was on average about 5 st higher than that of H* accents (6.9 st vs. 1.9 st, $p < .001$; see Table 32).

	Estimate	Lower bound	Upper bound	p
intercept (H*)	1.98	1.47	2.52	0.0001
L+H*	4.95	3.88	6.17	0.0001

Table 32: Excursion estimates (in st), lower and upper bounds and probability levels for H* and L+H* accents followed by a low boundary tone. N = 68.

Similarly, L+H* had a SLOPE that was about 12.7 st/sec steeper than that of H* accents (21.7 st/sec vs. 8.9 st/sec, $p < .001$; see Table 33).

	Estimate	Lower bound	Upper bound	p
intercept (H*)	8.93	5.46	12.24	0.0001
L+H*	12.77	8.22	17.05	0.0001

Table 33: Slope estimates (in st/sec), lower and upper bounds and probability levels for H* and L+H* accents followed by a low boundary tone. N = 68.

The same differences were found in cases where these accents were followed by a high boundary tone. EXC in L+H* accents was on average 6.1 st higher than in H* accents (8.91 vs. 2.74 st/sec, $p < .001$; see Table 34), and SLOPE was about 14.5 st/sec steeper than in H* accents (27.1 st/sec vs. 12.6 st/sec, $p < 0.001$; see Table 35).

Appendix B Pitch accent label consistency check

	Estimate	Lower bound	Upper bound	p
intercept (H*)	2.74	1.37	4.28	0.0004
L+H*	6.17	4.19	8.10	0.0001

Table 34: Excursion estimates (in st), lower and upper bounds for H* and L+H* accents followed by a high boundary tone. N = 42.

	Estimate	Lower bound	Upper bound	p
intercept (H*)	12.64	6.48	17.64	0.0001
L+H*	14.53	7.46	22.50	0.0002

Table 35: Slope estimates (in st/sec), lower and upper bounds for H* and L+H* accents followed by a high boundary tone. N = 42.

L+H* vs. L*+H

There is only one occurrence of a L*+H accent with a low boundary tone, which makes a statistical comparison with L+H* accents impossible. In this one instance of L*+H, the pitch maximum occurs about 31% later (102.2% into the accented syllable) than in the average L+H* accent that is followed by a low boundary tone (72.4% into the accented syllable). The same holds for the pitch minimum, which occurs also later in this L*+H accent (11.2% vs. 1.4%).

When followed by a high boundary tone, the pitch minima in L+H* accents are aligned 27.9% earlier than in L*+H accents (1.8% into the accented syllable vs. 27.6%, $p < .001$; see Table 36).

Appendix B Pitch accent label consistency check

	Estimate	Lower bound	Upper bound	p
intercept (L+H*)	-0.29	-5.81	4.77	0.87
L*+H	27.96	22.20	35.25	0.0001

Table 36: Estimates for the position of the pitch minimum, upper and lower bounds, for L+H* and L*+H accents, followed by a high boundary tone. N = 71.

L* vs. L*+H

Since there were no instances of L* occurring with a low boundary tone, the comparison with L*+H is done only for accents followed by a high boundary tone. L*+H accents have a significant higher EXC than L* accents (6.3 st/sec vs. 4.8 st/sec, $p < .05$; see Table 37), but the two accents do not differ significantly from each other in SLOPE (see Table 38). However, there was a significant difference in temporal alignment: The pitch minimum occurred later in L* accents compared to L*+H accents (43.1% into the accented syllable vs. 28.6%, $p < .001$, see Table 39).

	Estimate	Lower bound	Upper bound	p
intercept (L*+H)	6.36	-0.81	13.70	0.0696
L*	-1.56	-2.91	-0.05	0.0332

Table 37: Excursion estimates, upper and lower bounds for L* and L*+H accents followed by a high boundary tone. N = 76.

Appendix B Pitch accent label consistency check

	Estimate	Lower bound	Upper bound	p
intercept (L*+H)	26.30	10.30	41.67	0.014
L*	-5.23	-12.04	1.72	0.126

Table 38: Slope estimates, upper and lower bounds for L* and L*+H accents followed by a high boundary tone. N = 76.

	Estimate	Lower bound	Upper bound	p
intercept (L*)	28.76	18.71	38.26	0.0012
L*+H	14.41	6.04	22.57	0.0004

Table 39: Estimates for position of the pitch minimum, upper and lower bounds for L* and L*+H accents followed by a high boundary tone. N = 76.

The fact that the two accent types did not differ significantly from each other in terms of SLOPE and only at $p < .05$ for EXCURSION seems at first surprising. However, it needs to be pointed out that the set of analyzed words also contained 26 *kamel* (/ka'me:l/) items. In these words, the accent falls on the final syllable, which means that pitch accent and boundary tone are aligned with the same syllable. Differences in EXC or SLOPE would not be expected in this case. The *kamel* items could thus blur any differences between these two accent types. If the model is constructed again without these iambic items, both the EXC and the SLOPE analysis turn out to be highly significant different for the two accent types. EXC in L*+H accents was significantly higher than in L* accents (4.4 st vs. 1.9 st, $p < .001$; see Table 40), and SLOPE in L*+H accents was significantly steeper than in L* accents (25.9 st/sec vs. 11.9 st/sec, $p < 0.001$; see Table 41).

Appendix B Pitch accent label consistency check

	Estimate	Lower bound	Upper bound	p
intercept (L*+H)	4.47	3.95	5.08	0.0001
L*	-2.56	-3.47	-1.83	0.0001

Table 40: Excursion estimates, upper and lower bounds for L* and L*+H accents followed by a high boundary tone, excluding *kamel* items. N = 50.

	Estimate	Lower bound	Upper bound	p
intercept (L*+H)	25.91	21.96	29.65	0.0001
L*	-14.62	-21.51	-10.14	0.0001

Table 41: Slope estimates, upper and lower bounds for L* and L*+H accents followed by a high boundary tone, excluding *kamel* items. N = 50.

Excluding the *kamel* items leaves, however, open the question as to whether there was a significant difference between the two accent types in these items. Since EXC and SLOPE are not suitable to test this, I looked at the alignment of the F0 minimum instead. In L*+H accents, the trough is usually located in the stressed syllable, but it may be pushed to the left if there is tonal crowding. Tonal crowding occurs when several tones are associated with the same segmental material. This can lead to adjustments in the fine phonetic detail of the realization of an accent (Arvaniti et al., 2006a, p. 670), such as small differences in alignment. In the *kamel* items, three tones are associated with the final syllable: The nuclear tone (L*), the high trailing tone (+H) and the high boundary tone (H-%). This should have the consequence that the F0 minimum occurs earlier. In contrast, in L* accents, there are only two tones associated with the final syllable (L* and H-%), resulting in less tonal crowding, so that we would expect the F0 minimum to be located clearly within the stressed syllable. Indeed, the predictions are borne out by the data: The trough occurs

Appendix B Pitch accent label consistency check

significantly later in L* accents compared to L*+H accents (30.09% into the stressed syllable vs. 17.7% into the stressed syllable, $p < .01$, see Table 42).

	Estimate	Lower bound	Upper bound	p
intercept (L*+H)	17.77	12.802	23.31	0.0001
L*	12.32	4.474	21.88	0.0056

Table 42: Estimates for the position of the pitch minimum, upper and lower bounds for L* and L*+H accents followed by a high boundary tone, for *kamel* items only. N = 26.

H+!H* vs. H+L*

There was only one case of a H+!H* accent that was followed by a high boundary (here the pitch did not fall significantly further after the stressed syllable, but stayed level), and there were no combinations of H+L* with a high boundary tone. The analysis therefore looks only at the low boundary tone cases. Here, the EXC in H+!H* accents turned out to be on average 4 st lower than in H+L* accents (2.3 st vs. 6.8, $p < .001$; see Table 43).⁷⁷ H+!H* accents also had a SLOPE which was on average 22.7 st/sec less steep than that in H+L* accents (-8.9 st/sec vs. -31.7 st/sec, $p < .001$; see Table 44).

⁷⁷ The excursion value for H+L* accents (6.8 st) is comparable to the manipulated value in Baumann and Grice' (2006) perception study. There, a H+L* accent was characterized by a high of 240 Hz on the immediately preceding syllable, followed by a fall into a low of 170 Hz in the accented syllable. This is equivalent to an excursion of about 5.9 st.

Appendix B Pitch accent label consistency check

	Estimate	Lower bound	Upper bound	p
intercept (H+L*)	6.89	5.88	8.29	0.0001
H+!H*	-4.68	-6.27	-3.58	0.0001

Table 43: Excursion estimates, upper and lower bounds for H+!H* and H+L* accents followed by a low boundary tone. N = 44.

	Estimate	Lower bound	Upper bound	p
intercept (H+L*)	-31.74	-37.55	-25.97	0.0001
H+!H*	22.78	15.99	28.71	0.0001

Table 44: Slope estimates, upper and lower bounds for H+!H* and H+L* accents followed by a low boundary tone. N = 44.

H* vs. !H*

Since there were no cases of !H* accents in combination with a high boundary tone, only cases of accents followed by a low boundary tone were analyzed. The analysis showed that the difference in peak height between the peak preceding the pitch accent and the maximum of the pitch accent itself was on average 7.2 st larger in !H* accents than in H* accents (10.8 st vs. 18.0 st, $p < .05$; see Table 45).

Appendix B Pitch accent label consistency check

	Estimate	Lower bound	Upper bound	p
intercept (H*)	10.86	8.99	13.67	0.0001
!H*	7.20	1.32	10.17	0.0116

Table 45: Estimates for peak height difference between preceding peak and pitch accent maximum, upper and lower bounds for !H* and H* accents followed by a low boundary tone. N = 52.

7 Discussion

The foregoing analyses have tested predicted phonetic differences between deaccented and accented items on the one hand, and certain pairs of potentially confusable accents on the other.

Concerning the first test, it was shown that in low-ending phrases, accented items differ from deaccented items in terms of slope. Accented items always had a steeper slope than deaccented items, as predicted. Most accents also had a higher excursion than deaccented items. The same result was obtained for utterances with high boundary tones, with the exception of H*. The non-significant differences are probably due to the fact that excursion and slope were measured rather conservatively in these accents (i.e., excluding potential pitch rises on the post-accentual syllable), whereas in deaccented items the entire phrase was taken as the measuring domain (including final pitch rises due to upstep). It is also possible that prominence differences between these accent types and deaccented items are better reflected in other acoustic parameters such as intensity.

With respect to the second test (contrasting potentially confusable pairs of accents), the analyses showed that all seven accent types differ from each other along a number of acoustic dimensions in the predicted way. This provides substantial evidence that the accents have been labeled consistently according to the GToBI system.

Appendix C Consistency check of pitch accent labels by age groups

C Consistency check of pitch accent labels by age groups

Table 46 gives an overview of the statistical differences between accent pairs by age group. The details of each individual analysis are provided after that.

accent pair	Predicted difference	5 yrs.	7 yrs.	adults
H* vs. L+H* (L-%)	higher excursion in L+H* steeper slope in L+H*	** **	*** ***	*** ***
H* vs. L+H* (H-%)	higher excursion in L+H* steeper slope in L+H*	* **	*** ***	** **
L+H* vs. L*+H	pitch min. aligned later in L*+H	*	***	***
L* vs. L*+H	higher excursion in L*+H steeper slope in L*+H pitch min. later in L* (<i>kamel</i> items)	* * **	*** ** ns (✓)	*** *** ns (✓)
H+L* vs. H+!H*	higher excursion in H+L* steeper slope in H+!H*	** ns (✓)	- -	*** **
H* vs. !H*	Larger pitch difference between max. and preceding peak in !H*	ns (✓)	*	ns (✓)
deacc. vs. H*	higher excursion in H* steeper slope in H*	ns **	ns **	ns (✓) *
deacc. vs. L+H*	higher excursion in L+H* steeper slope in L+H*	** ***	*** ***	*** ***
deacc. vs. H+!H*	higher excursion in H+!H* steeper slope in H+!H*	ns ns	*** ***	ns (✓) ***
deacc. vs. !H*	higher excursion in !H* steeper slope in !H*	* ***	ns ***	- -
deacc. vs. L*	higher excursion in L* steeper slope in L*	ns (✓) ns (✓)	ns (✓) ns (✓)	** **
deacc. vs. L*+H	higher excursion in L*+H steeper slope in L*+H*	ns (✓) ns (✓)	ns (✓) ns (✓)	** ***

Table 46: Overview of statistical differences between accent pairs by age group. Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$. The symbol "(✓)" indicates that numerically the effect was in the predicted direction. A hyphen (-) occurs where there were too few instances (< 3) in one of the contrasted categories.

Appendix C Consistency check of pitch accent labels by age groups

1 Five-year-olds

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	2.5138	1.8082	3.2400	0.0000
H*	-0.4149	-1.5404	0.7106	0.4564
L+H*	1.7199	0.5169	2.9340	0.0058
H+L*	5.9022	2.9283	8.6429	0.0001
!H*	-1.3909	-2.5831	0.3265	0.0146
H+!H*	-0.4294	-1.5460	0.5344	0.3964

Table 47: Excursion estimates (in st), lower and upper bounds and probability levels for all accentuation types followed by a low boundary tone. N = 58.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	-4.461	-7.368	-1.736	0.0027
H*	11.654	7.422	16.265	0.0000
L+H*	21.100	16.255	26.029	0.0000
H+L*	-12.548	-24.044	-1.414	0.0293
!H*	11.495	7.199	15.974	0.0000
H+!H*	-2.722	-6.678	1.469	0.1798

Table 48: Slope estimates (in st/sec), lower and upper bounds and probability levels for all accentuation types followed by a low boundary tone. N = 58.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	1.736	-8.999	11.88	0.7418
H*	2.417	-10.527	13.05	0.8946
L+H*	8.798	-3.345	20.26	0.1504
L*+H	7.933	-4.130	18.42	0.1660
L*	2.540	-9.077	13.95	0.6320

Table 49: Excursion estimates (in st), lower and upper bounds and probability levels for all accentuation types followed by a high boundary tone. N = 22.

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	3.928	-18.095	27.07	0.6976
H*	13.107	-15.737	35.34	0.4218
L+H*	16.576	-10.591	38.72	0.2188
L*+H	21.309	-3.328	43.96	0.0928
L*	7.593	-16.501	31.97	0.5094

Table 50: Slope estimates (in st), lower and upper bounds and probability levels for all accentuation types followed by a high boundary tone. N = 22.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (H*)	2.075	1.1299	3.069	0.0002
L+H*	2.142	0.8092	3.686	0.0054

Table 51: Excursion estimates (in st), lower and upper bounds and probability levels for H* and L+H* accents followed by a low boundary tone, 5 years. N = 18.

	Estimate	Lower bound	Upper bound	p
intercept (H*)	7.085	3.393	10.78	0.0011
L+H*	9.761	4.508	15.62	0.0021

Table 52: Slope estimates (in st/sec), lower and upper bounds and probability levels for H* and L+H* accents followed by a low boundary tone. N = 18.

	Estimate	Lower bound	Upper bound	p
intercept (H*)	5.635	-2.7129	10.77	0.2364
L+H*	2.832	0.3108	14.33	0.0358

Table 53: Excursion estimates (in st), lower and upper bounds and probability levels for H* and L+H* accents followed by a high boundary tone. N = 8.

	Estimate	Lower bound	Upper bound	p
intercept (H*)	7.085	3.393	10.78	0.0014
L+H*	9.761	4.508	15.62	0.0014

Table 54: Slope estimates (in st/sec), lower and upper bounds and probability levels for H* and L+H* accents followed by a high boundary tone. N = 8.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (H+L*)	8.415	5.573	10.96	0.0002
H+!H*	-6.413	-9.149	-3.62	0.0006

Table 55: Excursion estimates (in st), upper and lower bounds for H+L* and H+!H* accents followed by a low boundary tone. N = 15.

	Estimate	Lower bound	Upper bound	p
intercept (H+L*)	-17.099	-31.562	-4.198	0.0214
H+!H*	9.952	-4.290	23.966	0.1500

Table 56: Slope estimates (in st/sec), upper and lower bounds for H+L* and H+!H* accents followed by a low boundary tone. N = 15.

	Estimate	Lower bound	Upper bound	p
intercept (L+H*)	1.238	-19.066	15.08	0.8876
L*+H	21.093	1.058	40.62	0.0282

Table 57: Estimates for the position of the F0 minimum, lower and upper bounds and probability levels for L+H* and L*+H accents followed by a high boundary tone. N = 12.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (L*+H)	5.973	2.652	8.7466	0.0074
L*	-4.250	-7.878	-0.5545	0.0338

Table 58: Excursion estimates (in st), lower and upper bounds and probability levels for L* and L*+H accents followed by a high boundary tone, excluding *kamel* items. N = 5.

	Estimate	Lower bound	Upper bound	p
intercept (L*+H)	22.06	13.09	30.142	0.0024
L*	-15.47	-24.56	-3.483	0.0210

Table 59: Slope estimates (in st/sec), lower and upper bounds and probability levels for L* and L*+H accents followed by a high boundary tone, excluding *kamel* items. N = 5.

	Estimate	Lower bound	Upper bound	p
intercept (L*+H)	11.82	-3.902	28.88	0.0722
L*	29.95	10.731	48.25	0.0082

Table 60: Estimates for the position of the F0 minimum, lower and upper bounds and probability levels for L* and L*+H accents followed by a high boundary tone, *kamel* items only. N = 8.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (H*)	10.812	6.268	15.87	0.0002
!H*	5.662	-1.414	12.20	0.1020

Table 61: Estimates for peak height difference between preceding peak and pitch accent maximum, upper and lower bounds for !H* and H* accents followed by a low boundary tone. N = 20.

2 Seven-year-olds

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	1.5292	1.0113	2.0688	0.0002
H*	-0.1082	-0.7269	0.4705	0.6548
L+H*	1.6980	1.0314	2.4105	0.0001
!H*	-0.0777	-0.9336	0.8929	0.9668
H+!H*	0.4387	-0.3584	0.9808	0.3496

Table 62: Excursion estimates (in st), lower and upper bounds and probability levels for all accentuation types followed by a low boundary tone. N= 59.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	0.4907	-2.505	3.423	0.7442
H*	6.1479	2.608	9.591	0.0012
L+H*	12.8778	8.844	16.955	0.0001
!H*	12.1705	6.982	17.782	0.0001
H+!H*	-8.1596	-12.082	-4.259	0.0001

Table 63: Slope estimates (in st/sec), lower and upper bounds and probability levels for all accentuation types followed by a low boundary tone. N= 59.

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	3.943	0.9787	6.796	0.0098
H*	-1.791	-5.6860	2.547	0.4272
L+H*	4.076	-0.1016	8.348	0.0456
L*+H	2.785	-0.6269	6.623	0.0996
L*	1.023	-2.7369	5.125	0.5510

Table 64: Excursion estimates (in st), lower and upper bounds and probability levels for all accentuation types followed by a high boundary tone. N= 54.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	8.7052	0.6664	15.127	0.0364
H*	0.1257	-9.1449	11.199	0.7896
L+H*	12.3461	2.4248	23.935	0.0102
L*+H	14.0498	7.0095	25.089	0.0006
L*	10.0467	2.0664	22.245	0.0230

Table 65: Slope estimates (in st/sec), lower and upper bounds and probability levels for all accentuation types followed by a high boundary tone. N= 54.

	Estimate	Lower bound	Upper bound	p
intercept (H*)	1.377	0.9792	1.771	0.0001
L+H*	1.870	1.2152	2.530	0.0001

Table 66: Excursion estimates (in st), lower and upper bounds and probability levels for H* and L+H* accents followed by a low boundary tone, 5 years. N = 34.

	Estimate	Lower bound	Upper bound	p
intercept (H*)	6.486	4.728	8.533	0.0001
L+H*	8.832	3.856	10.739	0.0001

Table 67: Slope estimates (in st/sec), lower and upper bounds and probability levels for H* and L+H* accents followed by a low boundary tone. N = 34.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (H*)	1.703	0.5919	3.715	0.0104
L+H*	6.634	3.7603	8.187	0.0001

Table 68: Excursion estimates (in st), lower and upper bounds and probability levels for H* and L+H* accents followed by a high boundary tone. N = 17.

	Estimate	Lower bound	Upper bound	p
intercept (H*)	7.649	3.883	13.81	0.0016
L+H*	15.290	5.974	20.26	0.0006

Table 69: Slope estimates (in st/sec), lower and upper bounds and probability levels for H* and L+H* accents followed by a high boundary tone. N = 17.

	Estimate	Lower bound	Upper bound	p
intercept (H+L*)	6.673	5.711	8.908	0.0001
H+!H*	-3.579	-6.940	-2.414	0.0001

Table 70: Excursion estimates (in st), upper and lower bounds for H+L* and H+!H* accents followed by a low boundary tone. N = 17.

	Estimate	Lower bound	Upper bound	p
intercept (H+L*)	-33.18	-41.079	-24.62	0.0001
H+!H*	17.59	5.829	29.81	0.0086

Table 71: Slope estimates (in st/sec), upper and lower bounds for H+L* and H+!H* accents followed by a low boundary tone. N = 17.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (L+H*)	0.1328	-9.384	9.86	0.9764
L*+H	30.4385	18.883	42.26	0.0001

Table 72: Estimates for the position of the F0 minimum, lower and upper bounds and probability levels for L+H* and L*+H accents followed by a high boundary tone. N = 25.

	Estimate	Lower bound	Upper bound	p
intercept (L*+H)	4.082	3.241	4.8406	0.0001
L*	-2.212	-3.511	-0.9453	0.0030

Table 73: Excursion estimates (in st), lower and upper bounds and probability levels for L* and L*+H accents followed by a high boundary tone, excluding *kamel* items. N = 20.

	Estimate	Lower bound	Upper bound	p
intercept (L*+H)	18.90	15.66	22.69	0.0001
L*	-8.05	-14.23	-3.06	0.0036

Table 74: Slope estimates (in st/sec), lower and upper bounds and probability levels for L* and L*+H accents followed by a high boundary tone, excluding *kamel* items. N = 20.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (L*+H)	17.78	7.294	28.85	0.0096
L*	10.32	-7.054	26.33	0.1520

Table 75: Estimates for the position of the F0 minimum, lower and upper bounds and probability levels for L* and L*+H accents followed by a high boundary tone, *kamel* items only. N = 7.

	Estimate	Lower bound	Upper bound	p
intercept (H*)	11.27	7.575	15.90	0.0001
!H*	14.36	2.240	21.07	0.0192

Table 76: Estimates for peak height difference between preceding peak and pitch accent maximum, upper and lower bounds for !H* and H* accents followed by a low boundary tone. N = 20.

Appendix C Consistency check of pitch accent labels by age groups

3 Adults

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	1.8576	-0.9006	3.476	0.2368
H*	5.7446	-0.9558	4.845	0.2028
L+H*	6.1543	2.9545	8.445	0.0001
H+L*	-0.5028	3.4558	8.907	0.0002
!H*	1.3093	-4.4791	3.502	0.7940
H+!H*	1.8576	-1.3086	4.220	0.3474

Table 77: Excursion estimates (in st), lower and upper bounds and probability levels for all accentuation types followed by a low boundary tone. N = 41.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	0.8425	-13.568	5.6290	0.4198
H*	14.0723	8.401	33.3729	0.0012
L+H*	35.8289	28.625	52.3364	0.0001
H+L*	-31.6544	-40.575	-17.0152	0.0001
!H*	4.3428	-6.297	28.6516	0.2234
H+!H*	-18.1465	-25.631	-0.3958	0.0440

Table 78: Slope estimates (in st/sec), lower and upper bounds and probability levels for all accentuation types followed by a low boundary tone. N = 41.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	0.9191	-4.012	6.168	0.6744
H*	2.8103	-1.225	6.631	0.1592
L+H*	6.4204	2.796	9.714	0.0004
L*+H	5.5990	2.487	8.665	0.0006
L*	5.1391	1.711	8.636	0.0070

Table 79: Excursion estimates (in st), lower and upper bounds and probability levels for all accentuation types followed by a high boundary tone. N = 59.

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	-0.5061	-14.523	13.26	0.9936
H*	18.1765	-1.870	37.20	0.0738
L+H*	35.8880	17.723	52.72	0.0001
L*+H	35.2517	19.682	50.85	0.0001
L*	27.0886	9.191	42.83	0.0018

Table 80: Slope estimates (in st/sec), lower and upper bounds and probability levels for all accentuation types followed by a high boundary tone. N = 59.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (H*)	3.035	0.5608	5.463	0.0162
L+H*	3.885	0.7604	7.388	0.0302

Table 81: Excursion estimates (in st), lower and upper bounds and probability levels for H* and L+H* accents followed by a low boundary tone, adults. N = 16.

	Estimate	Lower bound	Upper bound	p
intercept (H*)	15.76	7.504	27.29	0.0036
L+H*	21.02	6.716	32.63	0.0056

Table 82: Slope estimates (in st/sec), lower and upper bounds and probability levels for H* and L+H* accents followed by a low boundary tone. N = 16.

	Estimate	Lower bound	Upper bound	p
intercept (H*)	3.859	1.147	7.263	0.0136
L+H*	5.098	1.205	8.686	0.0106

Table 83: Excursion estimates (in st), lower and upper bounds and probability levels for H* and L+H* accents followed by a high boundary tone. N = 17.

	Estimate	Lower bound	Upper bound	p
intercept (H*)	16.46	4.252	30.39	0.0142
L+H*	18.47	1.080	34.02	0.0350

Table 84: Slope estimates (in st/sec), lower and upper bounds and probability levels for H* and L+H* accents followed by a high boundary tone. N = 17.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Lower bound	Upper bound	p
intercept (L+H*)	1.589	-7.57	7.504	0.9428
L*+H	27.461	19.77	37.455	0.0001

Table 85: Estimates for the position of the F0 minimum, lower and upper bounds and probability levels for L+H* and L*+H accents followed by a high boundary tone. N = 34.

	Estimate	Lower bound	Upper bound	p
intercept (L*+H)	4.678	3.906	5.561	0.0001
L*	-2.717	-4.278	-1.545	0.0002

Table 86: Excursion estimates (in st), lower and upper bounds and probability levels for L* and L*+H accents followed by a high boundary tone, excluding *Kamel* items. N = 25.

	Estimate	Lower bound	Upper bound	p
intercept (L*+H)	31.80	25.84	37.76	0.0001
L*	-20.35	-30.66	-11.94	0.0002

Table 87: Slope estimates (in st/sec), lower and upper bounds and probability levels for L* and L*+H accents followed by a high boundary tone, excluding *Kamel* items. N = 25.

Appendix C Consistency check of pitch accent labels by age groups

	Estimate	Std. Error	t-value	p
intercept (L*+H)	23.009	4.171	5.516	0.000372
L*	4.343	6.917	0.628	0.545717

Table 88: Estimates for the position of the F0 minimum, standard error, t-values and probability levels for L* and L*+H accents followed by a high boundary tone, *Kamel* items only. N = 11.

Appendix D Statistical analysis for target words only

D Statistical analysis for target words only

For boundary tones, the same results are obtained as in the combined analysis: Five-year-olds produced significantly fewer utterances with high boundary tones than both adults ($\beta = -1.50, z = -3.36, p < .001$) and seven-year-olds ($\beta = -1.26, z = -2.84, p < .01$). The seven-year-olds, on the other hand, did not differ from the adults ($p > .5$).

Regarding pitch accent type, the distribution was again very similar to the one in the combined analysis. Table 89 gives an overview of the effects of STATUS and AGE.

accentuation type	differences by STATUS	differences by AGE
deaccented	new < accessible < given	5 years = 7 years = adults
L*+H	new = accessible > given	$\begin{array}{ccc} 5 \text{ years} & < & \text{adults} \\ & \diagdown \quad \diagup & \\ & = & \\ & \diagup \quad \diagdown & \\ & 7 \text{ years} & \end{array}$
H+L*	given = accessible = new	5 years = 7 years < adults
!H*	given = accessible = new	$\begin{array}{ccc} 5 \text{ years} & > & \text{adults} \\ & \diagdown \quad \diagup & \\ & = & \\ & \diagup \quad \diagdown & \\ & 7 \text{ years} & \end{array}$
L*	$\begin{array}{ccc} \text{new} & > & \text{given} \\ & \diagdown \quad \diagup & \\ & = & \\ & \diagup \quad \diagdown & \\ & \text{accessible} & \end{array}$	5 years = 7 years = adults
L+H*	new > given new > accessible accessible > new > given new adults > accessible	for 5 years = 7 years for 7 years for adults 5 years = accessible 7 years

Table 89: Differences of pitch accent type distribution by information status and age for target words only. Statistically significant differences are highlighted in boldface.

Appendix E Within-accent category comparison across age groups

E Within-accent category comparison across age groups

H*

	Estimate	Lower bound	Upper bound	p
intercept (adults)	3.132	2.508	3.8716	0.0001
7 years	-1.038	-1.951	-0.1609	0.0258
5 years	-1.729	-2.563	-0.9668	0.0002

Table 90: Excursion estimates, upper and lower bounds and probabilities for H* followed by a low boundary tone. N = 40.

	Estimate	Lower bound	Upper bound	p
intercept (5 years)	2.0940	1.507	2.635	0.0001
7 years	-0.6914	-1.348	-0.005	0.0426
adults	1.0379	0.185	1.935	0.0188

Table 91: Excursion estimates, upper and lower bounds and probabilities for H* followed by a low boundary tone. N = 40.

	Estimate	Lower bound	Upper bound	p
intercept (adults)	17.65	14.88	20.138	0.0001
7 years	-10.49	-13.78	-7.036	0.0001
5 years	-10.69	-13.83	-7.837	0.0001

Table 92: Slope estimates, upper and lower bounds and probabilities for H* followed by a low boundary tone. N = 40.

Appendix E Within-accent category comparison across age groups

	Estimate	Lower bound	Upper bound	p
intercept (5 years)	7.1595	4.987	9.265	0.0001
7 years	-0.2027	-3.108	2.020	0.7356
adults	10.4867	7.136	13.879	0.0001

Table 93: Excursion estimates, upper and lower bounds and probabilities for H* followed by a low boundary tone. N = 40.

L+H*

	Estimate	Lower bound	Upper bound	p
intercept (adults)	6.914	5.065	8.7405	0.0001
7 years	-2.757	-5.391	-0.1814	0.0404
5 years	-3.685	-6.151	-1.2236	0.0050

Table 94: Excursion estimates, upper and lower bounds and probabilities for L+H* followed by a low boundary tone. N = 28.

	Estimate	Lower bound	Upper bound	p
intercept (5 years)	4.157	2.3149	6.084	0.0006
7 years	-0.928	-3.4309	1.419	0.3914
adults	2.757	0.1645	5.340	0.0422

Table 95: Excursion estimates, upper and lower bounds and probabilities for L+H* followed by a low boundary tone. N = 28.

Appendix E Within-accent category comparison across age groups

	Estimate	Lower bound	Upper bound	p
intercept (adults)	35.89	28.48	43.699	0.0001
7 years	-19.37	-29.85	-7.907	0.0010
5 years	-22.39	-32.49	-12.299	0.0002

Table 96: Slope estimates, upper and lower bounds and probabilities for L+H* followed by a low boundary tone. N = 28.

	Estimate	Lower bound	Upper bound	p
intercept (5 years)	16.524	8.626	24.650	0.0004
7 years	-3.022	-14.029	6.781	0.5172
adults	19.367	8.758	30.655	0.0026

Table 97: Slope estimates, upper and lower bounds and probabilities for L+H* followed by a low boundary tone. N = 28.

	Estimate	Lower bound	Upper bound	p
intercept (adults)	-8.956	-19.170	0.7591	0.0680
7 years	15.447	1.812	30.0226	0.0334
5 years	15.849	2.917	29.3168	0.0170

Table 98: Estimates for the position of the pitch minimum, upper and lower bounds and probabilities for L+H* followed by a low boundary tone. N = 28.

Appendix E Within-accent category comparison across age groups

	Estimate	Lower bound	Upper bound	p
intercept (5 years)	6.4912	-4.387	16.0818	0.2344
7 years	0.4017	-13.154	14.0280	0.9188
adults	-15.447	-28.919	-0.6176	0.0330

Table 99: Estimates for the position of the pitch minimum, upper and lower bounds and probabilities for L+H* followed by a low boundary tone. N = 28.

	Estimate	Lower bound	Upper bound	p
intercept (adults)	85.41	70.84	100.227	0.0001
7 years	-23.64	-45.51	-3.010	0.0350
5 years	-16.57	-36.79	2.625	0.0876

Table 100: Estimates for the position of the pitch maximum, upper and lower bounds and probabilities for L+H* followed by a low boundary tone. N = 28.

	Estimate	Lower bound	Upper bound	p
intercept (5 years)	61.77	46.723	77.85	0.0001
7 years	7.07	-14.827	26.00	0.5446
adults	23.64	2.219	44.17	0.0318

Table 101: Estimates for the position of the pitch maximum, upper and lower bounds and probabilities for L+H* followed by a low boundary tone. N = 28.

Appendix E Within-accent category comparison across age groups

L*+H

	Estimate	Lower bound	Upper bound	p
intercept (adults)	6.4988	2.907	9.815	0.0068
7 years	-1.5309	-4.582	2.296	0.4404
5 years	0.0045	-2.817	2.760	0.9806

Table 102: Excursion estimates, upper and lower bounds and probabilities for L*+H, followed by a high boundary tone. N = 48.

	Estimate	Lower bound	Upper bound	p
intercept (5 years)	4.968	1.108	9.399	0.0198
7 years	1.535	-2.052	4.641	0.4442
adults	1.531	-2.386	4.556	0.4430

Table 103: Excursion estimates, upper and lower bounds and probabilities for L*+H, followed by a high boundary tone. N = 48.

	Estimate	Lower bound	Upper bound	p
intercept (adults)	34.73	25.59	44.377	0.0008
7 years	-18.30	-28.07	-8.711	0.0010
5 years	-12.39	-20.58	-4.319	0.0056

Table 104: Slope estimates, upper and lower bounds and probabilities for L*+H, followed by a high boundary tone. N = 48.

Appendix E Within-accent category comparison across age groups

	Estimate	Lower bound	Upper bound	p
intercept (5 years)	16.427	5.858	28.00	0.0138
7 years	5.915	-3.717	14.92	0.2482
adults	18.303	8.352	27.45	0.0004

Table 105: Slope estimates, upper and lower bounds and probabilities for L*+H, followed by a high boundary tone. N = 48.

	Estimate	Lower bound	Upper bound	p
intercept (adults)	29.6906	19.46	39.004	0.0004
7 years	-5.7951	-19.63	6.927	0.3154
5 years	-0.0266	-10.59	9.963	0.9858

Table 106: Estimates for the position of the pitch minimum, upper and lower bounds and probabilities for L*+H, followed by a high boundary tone. N = 48.

	Estimate	Lower bound	Upper bound	p
intercept (5 years)	23.895	9.195	36.61	0.0014
7 years	5.769	-6.319	20.96	0.3482
adults	5.795	-6.898	19.52	0.3120

Table 107: Estimates for the position of the pitch minimum, upper and lower bounds and probabilities for L*+H, followed by a high boundary tone. N = 48.

Appendix E Within-accent category comparison across age groups

H+!H*

	Estimate	Lower bound	Upper bound	p
intercept (adults)	1.7779	1.0823	2.536	0.0001
7 years	0.2374	-0.7207	1.209	0.6164
5 years	0.9101	-0.3075	1.965	0.1368

Table 108: Excursion estimates, upper and lower bounds and probabilities for H+!H* accents, followed by a low boundary tone. N = 34.

	Estimate	Lower bound	Upper bound	p
intercept (5 years)	2.0154	1.4125	2.6892	0.0001
7 years	-0.2374	-1.2141	0.6647	0.6162
adults	0.6727	-0.5212	1.6327	0.2958

Table 109: Excursion estimates, upper and lower bounds and probabilities for H+!H* accents, followed by a low boundary tone. N = 34.

F Texts used in the reading experiment

Biber given

Picture 1: Ein Affe steht am Ufer eines Flusses, und hat ein Spielzeugboot in den Händen.

Picture 2: Dann setzt er das Boot aufs Wasser, und lässt es schwimmen.

Picture 3: Aber die Strömung ist sehr stark, und das Boot geht unter.

Picture 4: Da weint der Affe.

Picture 5: Doch ein Biber hat das Boot gerettet, und gibt es dem Affen zurück.

Picture 6: Der Affe freut sich, und umarmt den Biber.

Biber new/accessible

Picture 1: Ein Mann spaziert durch den Wald.

Picture 2: Er kommt an einen kleinen See, und streichelt einen Biber.

Picture 3: Dann geht er weiter durch den Wald.

Picture 4: An einer Lichtung pflückt er ein paar Blumen.

Picture 5: Da trifft er eine Bekannte und winkt ihr zu. Sie winkt zurück.

Picture 6: Er schenkt ihr die Blumen.

Picture 7: Dann geht er wieder alleine zurück.

Picture 8: Er gelangt wieder an den See, und nimmt den Biber mit.

Biene given

Picture 1: Eine junge Frau sitzt im Garten am Tisch und möchte frühstücken.

Picture 2: Sie macht sich ein Brot mit Honig.

Picture 3: Da setzt sich plötzlich eine Biene auf das Brot.

Picture 4: Die Frau ist sauer, und erschlägt die Biene.

Picture 5: Dabei erwischt sie das Honigbrot, das Brot klebt an ihrer Hand.

Appendix F Texts used in the reading experiment

Biene new/accessible

Picture 1: ein Künstler steht in seinem Atelier vor einer leeren Leinwand.

Picture 2: Er nimmt einen Pinsel, und malt eine Biene.

Picture 3: Dann verlässt er das Zimmer und geht in die Küche.

Picture 4: Dort sitzt eine Frau am Tisch, und trinkt Kaffee.

Picture 5: Die Frau gibt dem Künstler einen Apfel.

Picture 6: Der Künstler isst den Apfel.

Picture 7: Danach geht er wieder zurück in sein Atelier.

Picture 8: Er nimmt eine schere, und schneidet die Biene aus.

Kamel given

Picture 1: Ein kleines Mädchen geht in den Zoo.

Picture 2: An einem Verkaufsstand kauft es eine Packung Tierfutter.

Picture 3: Es kommt zu einem Gehege, das aber leer zu sein scheint.

Picture 4: Es wirft ein bisschen von dem Tierfutter in das Gehege.

Picture 5: Da zeigt sich ein Kamel.

Picture 6: Das Mädchen nimmt seine Kamera, und fotografiert das Kamel.

Picture 7: Nun kommen ein paar Vögel und picken dem Kamel das Futter weg.

Kamel new/accessible

Picture 1: Ein Mann in Safari-Ausrüstung verlässt sein Haus.

Picture 2: Er hat eine leere Flasche dabei.

Picture 3: Er macht sich auf den Weg, und reitet auf einem Kamel.

Picture 4: Dann geht er zu Fuß weiter, er ist auf der suche nach Wasser.

Picture 5: Er trifft einen Einheimischen, der ihm den weg zu einer Wasserstelle weist.

Picture 6: Er findet die Stelle, und füllt seine Flasche auf.

Picture 7: Dann macht er sich auf den Heimweg.

Picture 8: Er kommt wieder nach Hause, und streichelt das Kamel.

Appendix F Texts used in the reading experiment

Möwe given

Picture 1: Eine Robbe springt von einer Eisscholle ins Meer.

Picture 2: Unter Wasser fängt sie einen bunten Fisch.

Picture 3: Sie legt den Fisch neben sich auf der Eisscholle ab, und sonnt sich.

Picture 4: Da will eine Möwe den Fisch klauen.

Picture 5: Die Robbe ist wütend, und verjagt die Möwe.

Picture 6: Dann kommt eine andere Robbe, und will auch ein Stück von dem Fisch.

Möwe new/accessible

Picture 1: Eine Frau geht am strand spazieren, sie hat einen Beutel dabei.

Picture 2: Sie holt Körner aus dem Beutel, und füttert eine Möwe.

Picture 3: Dann läuft sie weiter den strand entlang.

Picture 4: Sie trifft eine Eisverkäuferin mit einem Eiswagen.

Picture 5: Sie kauft sich ein gemischtes Eis.

Picture 6: Die Frau isst das Eis, und geht langsam wieder zurück.

Picture 7: Sie hat das eis schon fast aufgegessen.

Picture 8: Dann holt sie ihre Kamera heraus, und fotografiert die Möwe.

Appendix G Pitch accent label consistency check for the reading experiment

G Pitch accent label consistency check for the reading experiment

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	1.8337	1.0320	2.708	0.0004
H*	0.9455	0.0321	1.797	0.0348
L+H*	4.2208	2.9008	5.584	0.0001
H+!H*	1.2132	0.4604	1.903	0.0016
H+L*	4.5255	3.8387	5.234	0.0001

Table 110: Excursion estimates (in st), lower and upper bounds and probability levels for all accentuation types. Since there was only one token of a L* accent and only two tokens of L*+H, these were not included. N = 82.

	Estimate	Lower bound	Upper bound	p
intercept (deacc.)	-5.959	-9.66	-2.315	0.0016
H*	18.799	11.29	26.432	0.0001
L+H*	42.036	31.12	52.635	0.0001
H+!H*	-10.002	-15.94	-4.419	0.0008
H+L*	-20.237	-26.10	-14.269	0.0001

Table 111: Slope estimates (in st/sec), lower and upper bounds and probability levels for all accentuation types. Since there was only one token of a L* accent and only two tokens of L*+H, these were not included. N = 82

Appendix G Pitch accent label consistency check for the reading experiment

	Estimate	Lower bound	Upper bound	p
intercept (H*)	2.549	1.086	3.959	0.0018
L+H*	4.732	2.156	7.525	0.0064

Table 112: Excursion estimates (in st), lower and upper bounds and probability levels for H* and L+H* accents followed by a low boundary tone. N = 13.

	Estimate	Lower bound	Upper bound	p
intercept (H*)	13.53	5.688	19.88	0.0022
L+H*	21.04	9.457	35.66	0.0022

Table 113: Slope estimates (in st/sec), lower and upper bounds and probability levels for H* and L+H* accents followed by a low boundary tone. N = 13.

	Estimate	Lower bound	Upper bound	p
intercept (H*)	3.073	1.901	3.926	0.0001
L+H*	3.096	2.299	4.820	0.0001

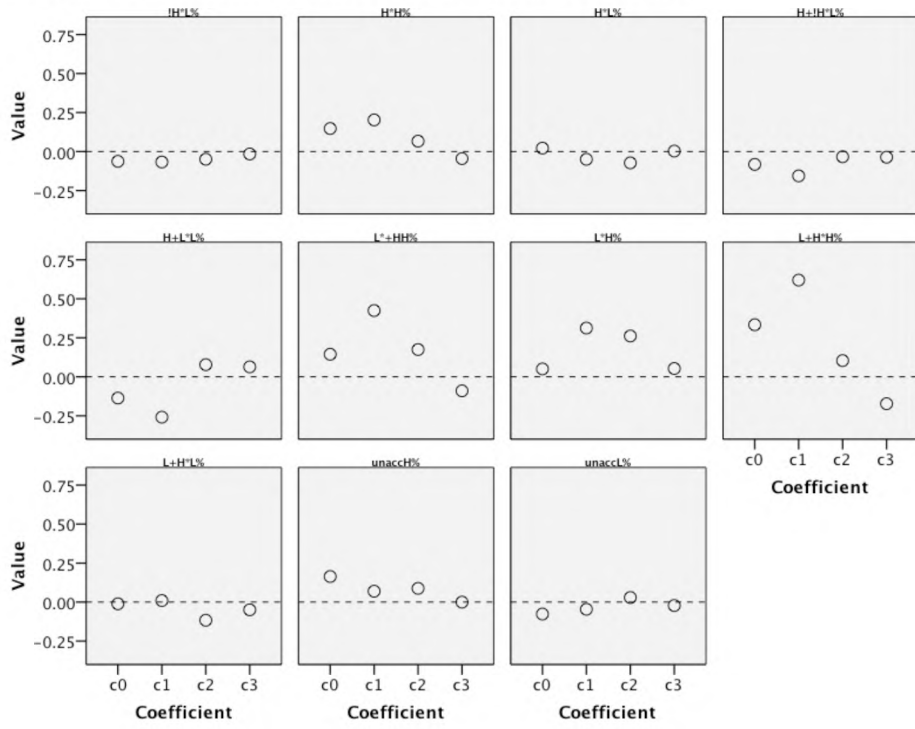
Table 114: Excursion estimates (in st), lower and upper bounds and probability levels for H+!H* and H+L* accents followed by a low boundary tone. N = 38.

	Estimate	Lower bound	Upper bound	p
intercept (H*)	-16.278	-21.26	-10.482	0.0001
L+H*	-9.617	-17.87	-2.436	0.0152

Table 115: Slope estimates (in st/sec), lower and upper bounds and probability levels for H+!H* and H+L* accents followed by a low boundary tone. N = 38.

Appendix H Four-coefficient profiles for nuclear contours

H Four-coefficient profiles for nuclear contours



Samenvatting

Als mensen met elkaar praten, refereren zij aan allerlei zaken, zoals personen (*Beatrix, de mevrouw met de grote hoed, zij*), plaatsen (*Amersfoort, links naast de HEMA, daar*) en objecten (*de Nachtwacht, Hans z'n fiets, het*) enzovoorts. De zaken (referenten) waarnaar wordt verwezen hoeven niet per sé aanwezig te zijn in het hier en nu. Sprekers hebben namelijk representaties van deze objecten en situaties opgeslagen in hun hoofd en als zij met elkaar in gesprek zijn kunnen zij daar toch naar verwijzen. Dit verwijzen gebeurt met behulp van verschillende soorten naamwoorden die variëren in hun morfosyntactische vorm en hun lexicale uitgebreidheid, zoals *eigennamen* ("Henk", "Marieke"), *bepaalde naamwoordfrases* ("de mevrouw met de grote hoed", "het huisje"), *voornaamwoorden* ("ik", "zij", "deze") of zelfs zogenaamde 'lege' elementen, zoals het weggelaten subject in de zin "Marieke deed haar best maar \emptyset faalde".

Er zijn verschillende factoren die bepalen welke vorm een spreker gebruikt. Bepaalde naamwoordfrases, zoals "het boek", veronderstellen bijvoorbeeld dat een object al bekend is bij de gesprekspartner. Behalve het verwijzwoord zelf kunnen sprekers ook nog de prosodische vorm ervan variëren: een en hetzelfde woord kan met verschillende toonhoogtecontouren (of 'toonhoogteaccenten') worden uitgesproken. Men kan bijvoorbeeld *Beatrix?* met een stijgende toonhoogte, of *Beatrix!* met een dalende toonhoogte uitspreken. In de literatuur wordt aangenomen dat de keuze voor een toonhoogteaccent onder andere wordt beïnvloed door de *informatiestatus* van een referent. Met informatiestatus wordt dan de veronderstelde aan- of afwezigheid van de referent in het (kortetermijn-)geheugen van de luisteraar bedoeld. Bij de aanwezigheid van een referent gaat het om de mate waarin die referent op het moment van spreken in het geheugen geactiveerd is. Als er een paar woorden eerder over een referent is gesproken, heeft hij namelijk een andere informatiestatus dan wanneer hij een half uur geleden ter sprake is gekomen. Behalve de informatiestatus van een referent kan ook de *pragmatische functie* van een referent in de zin de keuze voor een toonhoogteaccent beïnvloeden. Zo kan een referent het *topic* van de zin zijn. Dit betekent, dat de uiting over deze referent gaat en de spreker er in zijn uiting informatie aan toevoegt (bijv. *Beatrix heeft een*

Samenvatting

hoed op). Een referent kan ook de focus van een zin zijn, in welk geval de referent dat deel van de zin is wat de spreker tot de meest saillante informatie maakt (*Beatrix heeft een hoed op*). De factoren die bepalen welk soort referent een spreker kiest en welke prosodische vorm hij daarvoor gebruikt behoren tot wat in de taalwetenschap *informatiestructuur* genoemd wordt. Sprekers zijn zich ervan niet of nauwelijks bewust; maar voor kinderen die hun moedertaal leren is het een heel proces om te leren wanneer zij een bepaalde (in het Duits bijvoorbeeld: “der Hund”) of onbepaalde naamwoordzin (“ein Hund”) moeten gebruiken en met welke toonhoogteaccent zij deze moeten uitspreken.

In dit proefschrift wordt onderzocht hoe Duitse kinderen en volwassenen intonatie gebruiken om de informatiestatus en de topic-functie (of *topicaliteit*) van referenten uit te drukken. Ook bevat dit proefschrift enkele studies waarin een aantal methodologische aspecten van de analyse van kinder- en volwassenenprosodie (d.w.z., toonhoogte en ritme) wordt onderzocht.

Eerder onderzoek op het gebied van informatiestructuur en intonatie in het Duits was ofwel gebaseerd op voorgelezen taal (bij volwassenen) ofwel op imitatietaken (bij kinderen). De studies in dit proefschrift daarentegen zijn gebaseerd op vrije, dus niet van tevoren gespecificeerde, spraak. Om deze vrije spraak te verkrijgen werden beeldverhalen gebruikt. De proefpersonen kregen een serie afbeeldingen te zien die samen een verhaal vormden en er werd aan hen gevraagd om dit verhaal aan een luisteraar vertellen. De luisteraar kon de plaatjes niet zien. Dit was belangrijk omdat de sprekers er hierdoor niet van uit konden gaan dat een bepaalde referent bekend (in taalkundige terminologie *gegeven* of *given*) was bij de luisteraar. Behalve in het type taaldata dat is onderzocht, verschilt dit proefschrift van eerdere studies omdat er zowel fonologische als fonetische analyses (in plaats van alleen één van de twee) zijn gedaan.

In de eerste studie in hoofdstuk 3 wordt de prosodische realisatie onderzocht van topicale referenten in het zogenaamde *voorveld*, d.w.z. de positie voor het finiete werkwoord. Uit de resultaten blijkt dat volwassenen voor topicale referenten voornamelijk de stijgende toonhoogteaccenten L+H* en L*+H gebruiken. Deze bevinding komt overeen met eerdere resultaten (Braun,

2006; Mehloorn, 2001). Een nieuw resultaat in de huidige studie is dat er ook veel H* accenten en gedeaccentueerde referenten worden gebruikt. Kinderen van vijf en zeven jaar oud blijken bovendien dezelfde toonhoogteaccenten als volwassenen te gebruiken, alleen met andere onderlinge frequenties: de vijf- en zevenjarigen gebruiken vaker een H* accent dan de volwassenen, en de vijfjarigen gebruiken nauwelijks L*+H accenten, in tegenstelling tot de volwassenen. Er zijn bovendien verschillen tussen kinderen en volwassenen op fonetisch gebied. Vergeleken met de volwassen sprekers hadden de toonhoogteaccenten van de zevenjarigen een kleinere bandbreedte. Verder produceerden beide kindergroepen hun toonhoogteaccenten met een langzamere toonhoogteverandering (dus met een minder steile helling). Tenslotte produceerden zij het aan de piek voorafgaande toonhoogtedal vroeger dan de volwassenen, terwijl de positie van de toonhoogtepieken onderling niet verschilde. Een mogelijke verklaring voor de vroegere positionering van de toonhoogtedalen is dat de kinderen de stijgende beweging vroeger moesten beginnen om met een minder steile helling het hoge eindpunt toch nog op tijd te bereiken. Ondanks deze verschillen in de fonetische realisatie van accenten laat de studie zien dat kinderen van vijf en zeven jaar oud sterk lijken op volwassenen: ze gebruiken dezelfde toonhoogteaccenttypes en dat drukken topicale referenten niet op een fundamenteel andere manier uit.

Een duidelijk verschil tussen volwassenen en kinderen werd echter wel gevonden in *de mate* waarin topicale referenten in het voorveld voorkwamen. Bij de kinderen kwamen namelijk – anders dan bij de volwassenen – nauwelijks referenten in zinsinitiële positie voor. Zij vulden het voorveld vaak met voegwoorden en bijwoordelijke bepalingen van tijd zoals *und dann* ('en toen'). Dit komt overeen met resultaten uit eerdere studies naar vertellingen door kinderen (Hickmann et al., 1996). Hoewel deze bevinding niet direct relevant is voor de onderzoeksvraag van deze dissertatie, laat zij toch iets zien over de strategieën die sprekers van verschillende leeftijden gebruiken om een vertelling te structureren. Volwassenen nemen aan dat de tijd in een verhaal sequentieel verloopt en ze markeren dit niet noodzakelijkerwijs lexicaal. Zij besteden meer aandacht aan de globale thematische structuur van een vertelling, en bouwen hun verhalen op rondom de protagonisten en hun handelin-

Samenvatting

gen. Kinderen daarentegen leggen meer nadruk op de lokale temporele koppeling van de gebeurtenissen, die zij uitdrukken met combinaties van voegwoorden en bijwoordelijke bepalingen van tijd. Zoals hieronder wordt besproken, duiden ook de gegevens over de uitdrukking van referenten die aan het einde van een zin voorkomen (hoofdstuk 4) erop dat kinderen tijdsmarkeringen gebruiken om lokale gebeurtenissen aan elkaar te koppelen. Volwassenen gebruiken eerder intonatie om aan te geven dat hun vertelling nog niet afgelopen is.

In hoofdstuk 4 wordt de markering van de informatiestatus van frasenfinale referenten onderzocht, of om precies te zijn: de markering van *activatie*. De referenten in de beeldverhalen hadden steeds één van de volgende drie informatiestatussen: *nieuw* (nog nooit in het verhaal voorgekomen), *gegeven* (in het vooraangaand plaatje voorgekomen), of *toegankelijk* (een paar plaatjes geleden voorgekomen). De vraag in deze studie was of elke informatiestatus – zoals in de vakliteratuur wordt aangenomen – met een speciaal accent gemarkeerd wordt. Uit de resultaten voor de volwassenen blijkt dat de sprekers nieuwe referenten altijd accentueerden, maar er blijkt geen bewijs te zijn voor een typisch ‘nieuwheidsaccent’. Een onverwachte bevinding is dat L* en L*+H accenten vaker voor nieuwe referenten werden gebruikt dan voor gegeven referenten. Dit resultaat is verrassend, omdat Pierrehumbert en Hirschberg (1990) in een invloedrijke studie stelden dat toonhoogteaccenten die L* bevatten gegevenheid markeren. Het vaker voorkomen van L*-bevattende toonhoogteaccenten in de huidige studie zou een gevolg kunnen zijn van het veelvuldig gebruik van hoge grenstonen om continuering aan te geven. Sprekers lijken namelijk een voorkeur te hebben voor het combineren van lage toonhoogteaccenten met hoge grenstonen en andersom (Dainora, 2002b). De resultaten van hoofdstuk 4 laten verder zien dat een paar bekende generalisaties (bijv. ‘nieuwe referenten worden altijd met H* gemarkeerd’) niet kloppen als intonatie in een ruimere context wordt bekeken (in dit geval, in een context van zowel lage als hoge grenstonen). In tegenstelling tot nieuwe referenten werden gegeven referenten vaker gedeaccentueerd, hetgeen in overeenstemming is met enkele eerdere studies (bijv. Baumann, 2006; Baumann & Hadelich, 2003; Brown, 1983). Het feit dat meer dan de helft van alle gegeven

referenten wel werd geaccentueerd laat echter zien dat het hier slechts om een lichte tendens gaat (zie bijv. ook Braun & Chen, submitted).

Ook voor een speciaal toegankelijkheidsaccent H+L*, zoals Baumann en Grice (2006) hebben voorgesteld, werd geen bewijs gevonden. Voor toegankelijke referenten werden namelijk dezelfde accenten gebruikt als voor nieuwe referenten; het aantal gevallen van deaccentuatie voor toegankelijke referenten lag tussen de aantallen voor nieuwe en gegeven referenten in. Dit bevestigt Lambrecht's (1994) stelling dat er geen eenduidig fonologisch correlaat voor toegankelijkheid is, maar dat de spreker zelf kan bepalen of hij toegankelijke referenten wil behandelen als gegeven of als nieuwe referenten.

Al met al trekken de resultaten van hoofdstuk 4 het idee dat er een directe relatie bestaat tussen informatiestatus en accenttype, zoals wel in de literatuur wordt aangenomen (Baumann, 2006; Baumann & Grice, 2006; Chen et al., 2007; Pierrehumbert & Hirschberg, 1990), in twijfel. Het lijkt eerder zo te zijn dat het accentueren van nieuwe referenten verplicht is, maar sprekers relatief vrij zijn in hun realisatie van gegeven en toegankelijke referenten.

Wat de taalontwikkeling betreft, blijkt dat kinderen van vijf jaar intonatie al op een volwassen manier gebruiken om informatiestatus mee uit te drukken. Zij accentueren nieuwe referenten en weten dat er bij gegeven referenten (dus: referenten die al eerder benoemd zijn) deaccentuering is toegestaan. Deze resultaten komen overeen met wat er voor eenvoudige zinsparen is gerapporteerd (Wonnacott & Watson, 2008), en laten zien dat kinderen dezelfde kennis ook in complexere structuren zoals vertellingen kunnen toepassen. Het feit dat beide groepen kinderen gegeven en toegankelijke referenten verschillend behandelen toont bovendien aan dat kinderen net als volwassenen gevoelig zijn voor *recency of mention* (d.w.z., hoe lang geleden het is dat iets genoemd werd), en dat zij intonatie gebruiken om dat aan te geven.

Tegelijkertijd heeft de studie ook een paar aan leeftijd gerelateerde verschillen onthuld. Anders dan de zevenjarigen en de volwassenen, gebruiken de vijfjarigen maar weinig uitingen met een stijgende intonatie aan het eind van een prosodische frase. Een dergelijke intonatie kan gebruikt worden door de spreker om te signaleren dat hij van plan is om nog meer te gaan zeggen. Om dit resultaat te verklaren werd geopperd dat kinderen deze zogenaamde dis-

Samenvatting

course-structurende functie van intonatie pas later verwerven. Dit is vooral interessant in het licht van het resultaat uit hoofdstuk 3 dat kinderen lexicale middelen gebruiken om temporele continuïteit aan te geven. Het is dus voorstelbaar dat kinderen beginnen met een lexicale strategie om samenhang in hun betoog te brengen, en daar op latere leeftijd mee ophouden, om net als volwassenen intonatie te gaan gebruiken om aan te geven dat ze nog niet klaar zijn met spreken. Een ander leeftijdsverschil werd gevonden voor het H+L* accent, dat bij allebei de kindergroepen nauwelijks voorkwam. Gebaseerd op bevindingen van Dombrowski (2003) en Kohler (1991), die vonden dat H+L* een “wetende houding” (*attitude of knowingness*) kan aangeven, werd aangenomen dat de volwassenen dit accent op deze manier gebruikt hebben, en dat een dergelijk gebruik van intonatie om houdingen mee uit te drukken misschien pas later wordt verworven. Het is echter ook mogelijk dat H+L* deel uitmaakt van een bepaald register voor vertellingen, dat alleen door volwassenen wordt gebruikt.

Behalve fonologische analyses zijn er in hoofdstuk 4 ook fonetische analyses gedaan aan de accenten die aan het einde van een frase voorkomen. De data laten zien dat kinderen het stijgende gedeelte van accenten met een kleinere bandbreedte en een vlakkere helling realiseerden dan volwassenen. Dit komt overeen met de bevindingen voor zinsinitiële (prenucleaire) accenten, en duidt erop dat kinderen weliswaar in staat zijn om een aantal accenten te onderscheiden, maar individuele accenten ook met zeven jaar nog niet helemaal op dezelfde manier realiseren als volwassenen.

In het laatste hoofdstuk van deel II, hoofdstuk 5, wordt het vermogen onderzocht van kinderen om ‘identificeerbaarheid’ (d.w.z. of de referent voor de luisteraar bekend is of niet) en ‘activering’ te onderscheiden. Daarvoor zijn vertellingen van de kinderen geanalyseerd waarin twee bekende figuren, *der Sandmann* (‘Klaas Vaak’) en *der Weihnachtsmann* (‘De Kerstman’) werden geïntroduceerd. Deze referenten waren identificeerbaar voor zowel de spreker als de luisteraar vanwege *cultural common ground* (oftewel, een gemeenschappelijke culturele achtergrond), maar nieuw in de vertelling. Net als de volwassenen realiseerden de kinderen deze nog niet eerder gebruikte (*unused*) referenten met dezelfde accenttypes als zij voor de gloednieuwe (*brand-*

new) referenten in hoofdstuk 4 hadden gebruikt. Dit toont aan dat ze geleerd hebben dat identificeerbaarheid op basis van gemeenschappelijke culturele achtergrond geen invloed heeft op hoe de activering van een referent wordt gemarkeerd.

Samenvattend: de resultaten van de hoofdstukken 3 tot en met 5 wijzen erop dat vijf- en zevenjarige kinderen het grootste deel van het volwassen accentrepertoire verworven hebben en dat zij de verschillende accenten correct gebruiken om er informatiestructuurcategorieën mee te markeren. Wat zij nog moeten leren is het gebruik van intonatie voor discourse-structurende en paralinguïstische doelen en de precieze fonetische realisering van toonhoogteaccenten.

Deel III van dit proefschrift is gericht op een aantal centrale methodologische vraagstellingen. In de studie in hoofdstuk 6 is getoetst of het verschil tussen de bevindingen in hoofdstuk 4 en eerdere studies in de vakliteratuur te maken heeft met verschillen in spreekstijl. Met andere woorden: in hoeverre is het verschil toe te schrijven aan het feit dat er spraak uit spontaan geproduceerde vertellingen in plaats van voorgelezen spraak is geanalyseerd? Om deze vraag te beantwoorden is voorgelezen spraak geanalyseerd, die werd verkregen door een groep van volwassen sprekers uitgeschreven teksten te laten voorlezen bij de beeldverhalen uit hoofdstuk 4. Wat betreft de markering van informatiestatus repliceren de resultaten eerdere bevindingen uit Baumann's (2006) studie met voorgelezen spraak. Met name gegeven referenten werden doorgaans gedeaccentueerd, en toegankelijke referenten werden bij voorkeur gemarkeerd met H+L* accenten. Daarnaast werden veel nieuwe referenten ook met H+!H* en H+L* gerealiseerd. Deze resultaten werpen een nieuw licht op de onverwachte resultaten in hoofdstuk 4: De intonatie van spontane spraak is anders dan intonatie in voorgelezen spraak, en bij het gebruik van een voorleestaak worden dezelfde resultaten als die van Baumann verkregen. In hoofdstuk 6 wordt beargumenteerd dat dit verschil tussen voorgelezen en spontane spraak wellicht ontstaat doordat de verschillende functies van intonatie verschillende gewichten krijgen, afhankelijk van de communicatieve situatie. In spontane spraak kan *floor-holding* (het behouden van de sprekerrol) en het aangeven van houdingen belangrijker zijn dan het marke-

Samenvatting

ren van informatiestatus. Omgekeerd spelen vele paralinguïstische functies een kleinere rol in voorlezen, waardoor de sprekers de mogelijkheid hebben om intonatie te gebruiken voor het volledig representeren van de informatie-structuur in de tekst. Daarnaast zou de typische markering van informatie-structuur zoals die in de literatuur wordt voorgesteld een culturele vaardigheid kunnen zijn die in de loop van het leesonderwijs wordt aangeleerd. Deze bevindingen hebben belangrijke implicaties voor eerstetaalverwerving: Om te beginnen is het de vraag of de intonatie van voorgelezen spraak als direct uitgangspunt kan dienen voor een taalmodel voor kinderen. De bevindingen onderstrepen de dringende noodzaak om theorieën die op niet-spontane spraak gebaseerd zijn ook aan natuurlijke data te toetsen.

In het experiment dat in hoofdstuk 7 wordt beschreven wordt onderzocht in hoeverre 'onwetende luisteraars' prosodische grenzen kunnen identificeren. Er waren twee belangrijke motivaties voor deze studie. Ten eerste bleek het bij de analyse van de referenten aan het eind van een zin in hoofdstuk 4 af en toe moeilijk om te beslissen of een referent gevolgd werd door een frasengrens of niet. Ten tweede bleken er in de literatuur tegenstrijdige beweringen te zijn gedaan over het vermogen van kinderen om prosodische grenzen al dan niet te markeren. Als uit de studie zou blijken dat ongetrainde luisteraars de locatie van frasengrenzen goed kunnen determineren, zouden deze hun oordelen gebruikt kunnen worden om frasengrenzen te bepalen. Ook zou de mate van overeenkomst tussen naïeve beoordelaars als indicator kunnen dienen voor de betrouwbaarheid van de grensindicaties die kinderen produceren. In het experiment dat wordt besproken in hoofdstuk 7 luisterden volwassen proefpersonen naar uitingen uit het corpus (afgespeeld op de werkelijke snelheid) en gaven zij op een geschreven transcriptie aan waar zij frasengrenzen waarnamen. De resultaten laten een behoorlijke overeenstemming tussen beoordelaars zien voor alle leeftijdsgroepen die bovendien te vergelijken is met de overeenstemming die in andere studies werd gevonden (bijv. Mo et al., 2008). Dit suggereert dat de methode geschikt is als een aanvullende methode voor de analyse van spontane kinderspraak. Daarnaast leveren de resultaten bewijs voor de stelling dat kinderen duidelijke akoestische cues produceren om prosodische structuur aan te brengen in hun uitingen.

In het laatste hoofdstuk – hoofdstuk 8 – staat de vraag centraal hoe men kwantificeerbare ondersteuning kan vinden voor prosodische labels. In hoofdstuk 4 werd een controleprocedure gebruikt die liet zien dat de verschillende accenten ook fonetisch van elkaar verschilden, en de accenten consequent gelabeld waren. Deze procedure is echter zeer tijdrovend, en de lokalisering van akoestische markeringspunten zoals F0 maxima/minima is niet altijd eenvoudig en kan worden beïnvloed door subjectieve waarnemingseffecten. In hoofdstuk 8 werd daarom een alternatieve methode gebruikt: nucleaire contouren (d.w.z., de laatste toonhoogteaccenten in de frase en de daaropvolgende grenstonen) werden wiskundig gemodelleerd door middel van Legendre-polynomen, een methode die vergelijkbaar is met de benadering die Grabe, Kochanski en Coleman (2007) voor het Engels ontwikkelden. Statistische analyses lieten zien dat bijna alle contouren significant van elkaar verschilden in minstens één van de vier coëfficiënten (met uitzondering van één paar). Deze resultaten zijn in overeenstemming met de resultaten die verkregen werden met de andere procedure en bevestigen daarmee de nauwkeurigheid van de toegekende labels. Ook werd aangetoond hoe informatie over de locatie van tonale doelwaarden in verhouding tot de lettergreepstructuur kan worden afgeleid uit de polynoombeschrijvingen. Dit maakt deze benadering verenigbaar met meer taalkundige annotatiesystemen. Een vergelijking van verschillende leeftijdsgroepen binnen één toonhoogteaccentcategorie bevestigde bovendien de verschillen tussen kinderen en volwassenen die gevonden werden in hoofdstuk 4, zoals een lagere tweede coëfficiënt (helling) bij de kinderen. Sommige verschillen werden echter alleen met de polynoommethode gevonden, zoals een lagere derde coëfficiënt bij de kinderen dan bij de volwassenen.

Kortom, de methodologische studies in de hoofdstukken 5 tot en met 8 hebben twee hoofdbevindingen opgeleverd: Ten eerste kan de intonatie van voorgelezen spraak niet zonder meer als 'doelmodel' voor kinderen dienen. Ten tweede blijkt het zinvol om in prosodie-onderzoek bij kinderen meer methoden te gebruiken dan alleen puur impressionistische labels of geaggregeerde maten zoals gemiddelde toonhoogte. Oordelen van naïeve luisteraars kunnen experts informeren bij het annoteren. Wiskundige modellering kan

Samenvatting

kwantificeerbare akoestische gegevens opleveren die fonologische analyses kunnen ondersteunen en kunnen wezenlijk bijdragen aan nieuwe inzichten in de taalontwikkeling.

Dit proefschrift heeft een aantal resultaten opgeleverd die als aanknopingspunten kunnen dienen voor verder onderzoek. De vele overeenkomsten tussen kinderen en volwassenen laten zien dat vijfjarigen al heel vaardig zijn in het gebruiken van intonatie voor de structurering van hun uitingen. Op twee gebieden is echter nog duidelijk sprake van ontwikkeling: de fonetische 'fine-tuning' van toonhoogteaccenten en het gebruik van prosodie voor interactionele doelen. Toekomstig onderzoek zou zich kunnen richten op wanneer en hoe kinderen prosodie als een interactionele hulpbron inzetten.

Curriculum Vitae

Laura de Ruiter (née Herbst) was born in Berlin, Germany, on April 1st, 1980. After graduating from the Wald-Oberschule in Berlin in 1999, she studied General and Theoretical Linguistics at Potsdam University and Media and Communication Studies at the Freie Universität Berlin, where she also finished a one-year propaedeutics course in Japanese. A grant from the *Deutscher Akademischer Auslandsdienst* (DAAD) allowed her to subsequently spend several months as an intern in Tokyo, Japan, in 2001/2002. In 2004, she went to study at the University of Edinburgh, UK, where she obtained her Master's degree (with distinction) in Developmental Linguistics in 2005. In the same year, she was awarded a scholarship from the *Max-Planck-Gesellschaft zur Förderung der Wissenschaften* (MPG) to work on her doctoral dissertation at the Max Planck Institute for Psycholinguistics in Nijmegen, Netherlands. As a PhD student, Laura was PhD representative for the Humanities Section of the MPG PhDnet. Currently, she holds a post-doctoral position at the Cognitive Interaction Technology Excellence Cluster (CITEC) at Bielefeld University, Germany.

MPI Series in Psycholinguistics

1. *The electrophysiology of speaking. Investigations on the time course of semantic, syntactic and phonological processing.*
Miranda van Turenout
2. *The role of the syllable in speech production. Evidence from lexical statistics, metalinguistics, masked priming and electromagnetic midsagittal articulo-graphy.*
Niels O. Schiller
3. *Lexical access in the production of ellipsis and pronouns.*
Bernadette M. Schmitt
4. *The open-/closed-class distinction in spoken-word recognition.*
Alette Haveman
5. *The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach.*
Kay Behnke
6. *Gesture and speech production.*
Jan-Peter de Ruiter
7. *Comparative intonational phonology: English and German.*
Esther Grabe
8. *Finiteness in adult and child German.*
Ingeborg Lasser
9. *Language input for word discovery.*
Joost van de Weijer
10. *Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe.*
James Essegbey
11. *Producing past and plural inflections.*
Dirk Janssen
12. *Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea.*
Anna Margetts

MPI Series in Psycholinguistics

13. *From speech to words.*
Arie van der Lugt
14. *Simple and complex verbs in Jaminjung: A study of event categorization in an Australian language.*
Eva Schultze-Berndt
15. *Interpreting indefinites: An experimental study of children's language comprehension.*
Irene Krämer
16. *Language specific listening: The case of phonetic sequences.*
Andrea Weber
17. *Moving eyes and naming objects.*
Femke van der Meulen
18. *Analogy in morphology: The selection of linking elements in Dutch compounds.*
Andrea Krott
19. *Morphology in speech comprehension.*
Kerstin Mauth
20. *Morphological families in the mental lexicon.*
Nivja H. de Jong
21. *Fixed expressions and the production of idioms.*
Simone A. Sprenger
22. *The grammatical coding of postural semantics in Goemai.*
Birgit Hellwig
23. *Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies.*
Fermin Moscoso del Prado Martin
24. *Contextual influences on spoken-word processing.*
Daniëlle van den Brink
25. *Perceptual relevance of prevoicing in Dutch.*
Petra M. van Alphen

26. *Syllables in speech production: Effects of syllable preparation and syllable frequency.*
Joana Cholin
27. *Producing complex spoken numerals for time and space.*
Marjolein Meeuwissen
28. *Morphology in auditory lexical processing.*
Rachèl J.J.K. Kemps
29. *At the same time...: The expression of simultaneity in learner varieties.*
Barbara Schmiedtová
30. *A grammar of Jalonke argument structure.*
Friederike Lüpke
31. *Agrammatic comprehension: An electrophysiological approach.*
Marlies Wassenaar
32. *The structure and use of shape-based noun classes in Miraña (North West Amazon).*
Frank Seifart
33. *Prosodically-conditioned detail in the recognition of spoken words.*
Anne Pier Salverda
34. *Phonetic and lexical processing in a second language.*
Mirjam Broersma
35. *Retrieving semantic and syntactic word properties: ERP studies on the time course in language comprehension.*
Oliver Müller
36. *Lexically-guided perceptual learning in speech processing.*
Frank Eisner
37. *Sensitivity to detailed acoustic information in word recognition.*
Keren B. Shatzman
38. *The relationship between spoken word production and comprehension.*
Rebecca Özdemir
39. *Disfluency: Interrupting speech and gesture.*
Mandana Seyfeddinipur

MPI Series in Psycholinguistics

40. *The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation.*
Christiane Dietrich
41. *Cognitive cladistics and the relativity of spatial cognition.*
Daniel B.M. Haun
42. *The acquisition of auditory categories.*
Martijn Goudbeek
43. *Affix reduction in spoken Dutch.*
Mark Pluymaekers.
44. *Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence.*
Valesca Kooijman
45. *Space and iconicity in German Sign Language (DGS).*
Pamela Perniss
46. *On the production of morphologically complex words with special attention to effects of frequency.*
Heidrun Bien
47. *Crosslinguistic influence in first and second languages: Convergence in speech and gesture.*
Amanda Brown
48. *The acquisition of verb compounding in Mandarin Chinese.*
Jidong Chen
49. *Phoneme inventories and patterns of speech sound perception.*
Anita Wagner
50. *Lexical processing of morphologically complex words: An information-theoretical perspective.*
Victor Kuperman
51. *A grammar of Savosavo: A Papuan language of the Solomon Islands.*
Claudia Wegener
52. *Prosodic structure in speech perception and production.*
Claudia Kuzla

53. *The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension.*

Sarah Schimke

54. *Studies on intonation and information structure in child and adult German.*

Laura E. de Ruiter

Notes

Notes

Notes

Notes

Notes