

Intonation in Language Acquisition

Evidence from German

Inaugural – Dissertation
zur Erlangung des Doktorgrades
der Philosophischen Fakultät
der Universität zu Köln
im Fach Phonetik

Thomas Grünloh

Acknowledgements

First of all, I would like to thank my supervisors, Michael Tomasello & Elena Lieven from the Max Planck Institute for evolutionary Anthropology and Martine Grice from the University of Cologne, IfL – Phonetics. They not only gave me good advice, engaging discussions and supported me with my PhD research, they also gave me the freedom that I needed to find my own way.

To my family, I'd like to say thank you so much for always supporting me in whatever I've chosen to do.

So many people at the MPI-EVA have helped to make my work possible. In particular, I'd like to thank the nurseries, parents and children who took the time and effort to participate in my studies. Special thanks goes to Nadja Richter, Angela Loose and Manja Teich without whom testing wouldn't have been possible; also to Henriette Zeidler and Annett Witzmann who put so much effort into organizing trips, working life and dealing with administrative questions. I owe a great deal to Roger Mundry who helped make my statistics a breeze, to the research assistants in Leipzig, as well as Petra Jahn and her team.

Additionally, I wish to thank everyone at the Institut für Phonetik, Köln. I'm very grateful to all those institutes from whom I've received helpful comments at colloquia and help with administrative problems.

Additionally, I'd like to thank everybody in the Child Language group in Leipzig and Manchester for always being open to new (and sometimes crazy) research proposals. You've always supported my ideas, read my scripts - and joined me in celebrating Leipzig's night-life! You've all been an important part of my life for the last three years and, hopefully, you will continue to be so. Special thanks go to Sarah Girlich for being there when I needed your help with certain psychological questions, and Daniel Schmerse & Robert Hepach for those exciting kicker-games.

Also, I'd like to thank my folks in Cologne. I'm grateful also to Patrick and Sven for always offering me a corner on their couch, and to Silke and Lars for always being honest with me.

Finally I would like to thank Caro. You've made so many sacrifices in order for us to share our lives together. I'm grateful for every night and day you've watched over me. From the bottom of my heart, I thank you.

Abstract

This dissertation studies the role of intonation in language acquisition. After a general introduction about the phonetic and phonological aspects of intonation and its different forms and functions within language, two different models of language acquisition and the role of intonation within these two models will be presented.

Following this, I will present and discuss empirical data on the question, whether young German learning children use intonation in order to acquire language. Two comprehension studies will be presented. Here, I concentrate on the question whether children understand the referential function of intonation and whether they can use this knowledge in order to learn new words. Additionally, I will present empirical evidence that focuses on the question whether children use intonation in resolving participant roles in complex syntactic constructions as well as in resolving syntactic ambiguities development.

Finally, I will present two production studies that investigate the prosodic realization of target referents that have different informational statuses within a discourse from both young children and parents, talking to their children.

Overall, the data from these studies suggest that language learning children do use the intonational form of an utterance from early on in order to understand another's intention. Young language learning children do understand that a certain intonational form conveys a function. Additionally, the studies presented in this thesis suggest that children also use intonation in order to convey their own communicative intentions. Thus, intonation is an important instrument for young children's language acquisition as they use the information that is provided by intonation, not only to learn words and to combine them to syntactic constructions, but also for the understanding of paralinguistic properties of language.

The findings of the studies presented in this thesis are discussed with regard to different theories of language acquisition. Additionally, I will give insight into the understanding of the development of young children's use of intonation.

Contents

Acknowledgements.....	iii
Abstract.....	v
Part I: Theoretical Background.....	1
1. General introduction	3
2. Intonation	5
2.1. Introduction	5
2.2. The Phonetic aspects of intonation	6
Speech melody.....	6
Accentuation	7
2.3. The phonological aspects of intonation.....	10
2.3.1. Forms of intonation	10
Autosegmental and Metrical Phonology	10
GToBI.....	16
2.3.2. Functions of intonation.....	18
Affective functions	18
Intentional functions.....	21
2.4. Summary	28

3. Language Acquisition	31
3.1. The Nativist-Generative Approach.....	31
Bootstrapping mechanisms	35
3.2. Usage-Based Perspective	38
Intention reading	39
Pattern finding	41
3.3. The role of intonation in the two approaches	45
4. Intonation in language acquisition	48
4.1. Prerequisite.....	48
Perspective taking in infancy	49
Understanding communicative intentions	50
4.2. Intonation in Information Marking	51
5. Research questions	59
Part II: Empirical Studies - Comprehension	61
6. Referential function of intonation	63
6.1. Understanding intentions by intonation.....	63
6.1.1. Introduction	63
6.1.2. Data & Method	67
Participants	67
Materials and design.....	67
Procedure.....	68

Acoustic properties of the test material	71
Coding and reliability	73
6.1.3. Results & Discussion	73
6.2. Competition in Word Learning: Intonation vs. Mutual Exclusivity	76
6.2.1. Introduction	76
6.2.2. Data & Method	76
Participants	77
Materials, design, and procedure	77
Acoustic properties of the test material	78
Coding and reliability	79
6.2.3. Results & Discussion	80
6.3. General discussion	81
7. The role of intonation in grammatical constructions	83
7.1. Resolving syntactic ambiguities	83
7.1.1 Introduction	83
7.1.2. Data & Method	89
Participants	89
Materials and design	89
Acoustic properties of the test material	92
Procedure	94
Coding and Reliability	96
7.1.3. Results and Discussion	97
Children	97
Adult - control group	98

7.2. The role of context & intonation in resolving syntactic ambiguities	99
7.2.1. Introduction	99
7.2.2. Data & Method	100
Participants	100
Materials and design.....	100
Procedure.....	102
Coding and Reliability.....	103
7.2.3. Results and Discussion	103
7.3. General Discussion	106
Part III: Empirical Studies - Production	111
8. Young children’s intonational marking of new and given referents	113
8.1. Introduction	113
8.2. Data & Method	115
Participants	115
Materials	116
Design and Procedure	117
Coding and Reliability.....	118
Statistical Model for Main Analysis.....	122
8.3. Results and Discussion	123
Pitch accent type.....	123
Pitch range	124

9. The role of the input for children's intonational development	127
9.1. Introduction	127
9.2. Data & Method	128
Participants	128
Materials, Design and Procedure.....	128
Coding and Reliability.....	129
9.3. Results and Discussion	129
Pitch accent type.....	129
Pitch range	131
9.4. General Discussion	133
10. General discussion.....	136
10.1. Summary and Discussion of empirical findings.....	136
10.2. Open Questions and Future Research	141
10.3. Principal Conclusions	143
11. References	145
12. Appendix.....	163

Part I: Theoretical Background

1. General introduction

This dissertation studies the role of intonation in first language acquisition within the usage-based framework of language development (Tomasello, 2003). Within this framework, it is assumed that the process of language acquisition is based on diverse social-pragmatic and cognitive skills. Language is not seen as arising from an innate, modular system that follows linguistic principles and parameters (e.g. Chomsky 1980, 1993), but rather as an interplay between the overall cognitive abilities children need to understand others' communicative intentions and to communicate their own. Two sets of social and general cognitive skills are of particular importance: intention-reading and pattern-finding. Intention-reading skills allow prelinguistic infants, for example to share attention to events with others', establishing joint attentional frames and to understand others' communicative intentions. Additionally, pattern-finding skills are assumed to allow children to learn the structure of a language through using that language by means of powerful generalization abilities. Overall, the usage based approach assumes that it is the social-cognitive skills involved in reading and understanding the intentional and mental states of others that paves the way for language learning.

Research in the area of first language acquisition mainly focuses on the morpho-syntactic aspects of language. But, language consists of more than just a combination of morphemes and words into grammatical constructions. Within communication, it is not only important WHAT is said, but rather HOW it is said. The way an utterance is realized is mainly characterized by intonation. The intonational system fulfils a variety of different functions. It is active at many different levels of communication, in areas deemed purely linguistic, e.g. the division of utterances into informative and less informative parts, as well as areas considered more peripheral to linguistic inquiry, e.g. to signal emotional states of varying degrees of intensity, speaker affect, and attitude. What makes intonation so interesting for research into language acquisition is that a particular intonational form automatically conveys a certain function. For example, for West-Germanic languages (e.g. English, German and Dutch), it is typically assumed that information that is newly introduced within a discourse (and is thus important to the speaker) is marked with a pitch accent. On the other hand, information that is given (or less important) is characterized by the lack of an accent. This shows that the intonational realizations of utterances have a function - they convey the intention of a speaker, in this example what is important (or special and new) to him. However, in order to use the appropriate intonational form, a speaker has to know what is new or given in a situation – he needs the ability to understand what another person has in mind. And, in order to convey a certain function that fits with his communicative intention, the speaker has to use the appropriate intonational form. Reciprocally, the hearer also needs the knowledge about which

form conveys which function in order to understand the communicative intentions of a speaker.

In the current literature, it is not clear whether and/or when children do use intonation to understand others' intentions. But, this would seem to be an essential step because the intonational realization of utterances constitutes a great deal of the communicative intention. To understand and to learn a particular language, the child has to understand what another person is referring to and what that person intends to say: in other words, what that person has in mind. Intonation seems to be the perfect instrument in order to understand other people's intentions.

The studies presented in this thesis are intended to address research from two disciplines: that of developmental psychologists who are interested in the social-pragmatic and cognitive skills that are needed to acquire language; and that of phoneticians who are interested in young children's intonational development. My intention in addressing both psychologists and phoneticians is to bring these fields together. As language acquisition requires an understanding of others' intentions – an understanding that is centrally underpinned by the use of intonation - it seems that there should be more symbiosis between researchers of these fields in the study of language acquisition.

Since I am bringing together two partially intersecting fields of research, I shall first give separate accounts of their theoretical backgrounds in Part I of this thesis. In this introductory chapter, I will start by giving a broad overview of intonation, including its phonetic and phonological implementations. Additionally, I will provide an insight into the form – function mapping of intonation (Chapter 2.3.). Here, I will focus on both the affective function of intonation, in which intonation is produced subconsciously in speech, and the intentional functions of intonation, which are more under conscious control. Chapter 3 deals with different theories of language acquisition. Here, I will concentrate on two major theoretical frameworks, namely the Nativist-Generative account which assumes that children's capacity to acquire language depends on an "Universal Grammar", and the Usage-Based approach which assumes that the acquisition of language is based on overall social-pragmatic and cognitive skills. This background information is necessary in order to integrate the role of intonation in a theory of language acquisition. Following this, Chapter 4 will give an overview of the literature on infants' and children's ability to use intonation in the language acquisition process.

In the subsequent four chapters (Chapters 6 - 9), I will present empirical evidence investigating whether children can use intonation in order to understand others' intentions. First, I will focus on the question of whether children understand the intonational form of a request, based on whether or not the requested object was shared (Chapter 6.1.). Subsequently, Chapter 6.2. will deal with the question of what role intonation plays in the process of word-learning. Following this, Chapter 7 addresses the question of whether children can use intonation for the understanding of grammatical constructions. In Chapter 8, I will present an empirical study aimed at answering the question of how young children use intonation in order to realize the informational status of target

referents. Finally, in Chapter 9, I will consider the role that intonational input plays in the acquisition of intonation.

All these chapters start with a review of the literature in the specific field and finish with my empirical studies that are the heart of this thesis. Finally, in Chapter 10, I will finish with conclusions, theoretical speculations and some suggestions for future research.

2. Intonation

2.1. Introduction

When we hear someone on the street saying the word “Mary”, we hear a successive stream [m E ŋ i]. The meaning of a word is encoded in its phonological form. Beyond phonological form there are several other features intrinsic to spoken language that encodes meaning. Rather than providing information about **what** is spoken they give information about **how** it is said. Let us assume we hear an utterance like “This is Mary”. In written text without punctuation it is unclear what the speaker intends to say. In spoken language, in addition to the phonological meaning of the individual words a speaker has further ways to realize an utterance, because he can use a certain speech melody. For example, the sentence “This is Mary” can be uttered with a rising inflection at the end of the utterance. This would indicate that the speaker intends to ask whether the person in front of him really is Mary or not. Alternatively, a speaker could use a falling speech melody in order to make a statement and introduce Mary to another person. Features referring to this manner of speaking (including e.g. speech melody, pauses, amplitude) are known as the ‘suprasegmental’ features of language. The suprasegmental properties of speech play an important role in human communication. All spoken utterances require the presence of a voice. And, since the voice has physical and physiological implementations it is modulated at each point. This modulation of the voice and thus, the properties of the suprasegmental signal, may be expressed consciously or unconsciously. Thus, spoken language provides information about the intention and the emotional state of a speaker.

Speech is a complex communicative system, determined by linguistic, emotional and attitudinal factors. It provides diverse linguistic and paralinguistic functions with which a speaker can colour his utterance. These functions range from the marking of sentence¹ modality (question vs. statement) to the expression of emotional and attitudinal nuances (i.e. anger, fear, happiness).

¹ Following Sperber & Wilson (1995) I will use the term ‘sentence’ as referring to the purely linguistic properties (such as noun, pronoun and so on) and the term ‘utterance’ as including non-linguistic properties such as for example the discourse of utterances or the speaker’s intention.

Since the linguistic and paralinguistic features of language are all provided by the same cues, i.e. the physical and physiological properties of voice, which cannot be localized rigidly to particular segments, syllables, words or utterances, analyzing spoken language has proven a challenge to many researchers over recent decades. There have been many attempts to find one broad term to describe all of the features involved in spoken language. With respect to spoken language, the term 'intonation' is simply defined as the 'speech melody' or the 'pitch', meaning variations of the fundamental frequency (F0). But, the 'speech melody' of an utterance does not just contain the "ensemble of pitch variations in the course of an utterance" (t Hart et al. 1990: 10). It cannot be restricted to the movements of the fundamental frequency. For example, a rise in the speech melody automatically entails a longer duration of that movement (the higher the longer) and does not give any information about voice quality. A wider term was introduced to include all phenomena of the speech signal and its (para-) linguistic and physical correlates – 'prosody'. This definition of prosody covers all phenomena that are involved in the process of conveying a meaningful utterance, such as pitch movements and pitch range (speech melody or intonation), highlighting at word level (lexical stress) and utterance level (accentuation), the division of speech into chunks (phrasing), the marking of prominence relations (rhythm) and variations in speech rate (tempo). Not all of these prosodic components are included in abstract models of intonation at utterance level, but all may play a part in the signalling of discourse structure. Voice quality, for example, although often beyond the speaker's control (because of the influence of emotional state) can be modified for communicative purposes (e.g. intimacy).

This thesis focuses on young children's understanding of both the intentional and affective aspects of speech melody (intonation) as well as how (and why) certain parts of the speech stream can be made more salient than others. To understand how and why speech melody is as it is and what effect it has on both the speaker and the listener, I will explain the phonetic and the phonological implementations further.

2.2. The Phonetic aspects of intonation

Speech melody

The overall pattern of pitch movements within an utterance is what is commonly described as speech melody. It consists of more or less continuous, constantly changing pitch patterns. The pitch (or fundamental frequency – F0) is the prosodic feature that is most centrally involved in intonation. Physiologically, pitch is created by the vibrations of the vocal folds during the voiced parts of speech. It is primarily the result of muscular tension and the pressure of the air below and above the glottis and is dependent on the rate of vibrations of the vocal folds. This rate of vibration is reflected in the acoustic measurement of

fundamental frequency, measured in 'Hertz' (Hz). Hertz is defined as the unit of frequency i.e. the number of the cyclic opening and closing of the glottis per second. There are several determinates of the rate at which the vocal folds vibrate. Purely physiological determinates are their elasticity, length and mass. Variations in pitch are principally produced by the length and tension of the vocal cords, and these factors themselves are controlled by the intrinsic muscles of the larynx. Consequently, there are differences between genders, based on their body-size. For example, for males, the F0-range is typically between approximately 80 and 200 Hz, for females between approximately 180 and 400 Hz. For Young children, this range can be even higher. Another physiological influence, the pressure of air below the larynx, is commonly regarded as a secondary influence on the rate of vibration.

By actively controlling muscular tension and sub-glottal air pressure, a speaker has to a large extent active control over F0 (see Borden & Harris 1984: 74ff.). For example, she can produce rises and falls within the speech melody, or speak with high or low pitch. On the other hand, other physiological factors, cannot be actively controlled by the speaker, e.g. certain supralaryngeal articulatory gestures. Instead, these factors are influenced by unintended side-effects of vocalizations. For example, high vowels like /u/ and /i/ have higher intrinsic pitch than low vowels like /a/ (see e.g. Lehiste & Peterson, 1961; Ladd & Silverman, 1984). Additionally, a higher F0 at the beginning of a vowel is the result of the speech melody of a preceding voiceless obstruent (see Kingston, 1991; Gussenhoven 2004). These unintended aspects of speech produce minor interferences in the F0-pattern melody. However, although these interferences makes it difficult to identify the "original" speech melody, they do not influence listeners' interpretation of the intonation contour (see Silverman 1987) and are known as 'microprosody'.

Accentuation

Whereas the overall pattern of pitch movement is defined as the speech melody of an utterance, a single pitch movement associated with prominent syllables within that melody is commonly known as *accent*. Overall, both terms describe the relative emphasis that may be given to certain syllables in a word, or to certain words in a phrase or sentence. In the past, the word 'stress' and 'accent' have been used intertwined and in different and confusing ways. It has sometimes been used to describe prominence at word level, while other authors have used it to refer to prominence at the level of utterance. What both have in common is that prominences in terms of stress and accent have their productive and perceived bases in the physiological and physical properties of the speech organs. The following table (largely adopted from Baumann, 2006:12 & Uhmman 1991: 109) describes the phonetic parameters that constitute prominence in 'stress accent languages' like German and English and gives their correlates at the respective levels of description.

Table 1: phonetic parameters that generate accents and their correlates at different levels of description

Perception	Production	Acoustics
Pitch (High – Low)	quasi-periodic vibrations of the vocal folds	fundamental frequency (F0) in Hertz (Hz)
Loudness (loud – soft)	articulatory effort (e.g., air pressure)	Intensity in decibel (db)
Length (long – short)	articulation process	Duration in milliseconds (ms)
Vowel quality (full – reduced)	vocal tract configuration	spectral characteristics

Syllables that are in some sense stronger than other syllables, and are thus more prominent, have the potential to be described as stressed and accented. Which syllable is made stronger than others within a word is determined by language-specific rules for word-stress. In English or German, for example, the placement of prominence is not easily predictable. For this reason, the difference between strong and weak syllables is of some linguistic importance in these languages: in German, for example, the position of stress can change the meaning of a word ('UMfahren' - to knock down vs. 'umFAHren' - to drive around). The same is true for English e.g., 'IMport' (noun) and 'imPORT' (verb). Thus, prominence in terms of 'stress' forms part of the phonological composition of the word. At utterance level, some types of words typically occur in non-prominent form e.g., auxiliary verbs, pronouns, shorter prepositions or conjunctions. Other types of

words like nouns or main verbs are more likely to occur with prominence². Cruttenden (1986) assumes four different degrees of prominence (for English), depending on the effort that is put into its realization. 'Unstressed syllables' do not convey any prominence at all. 'Tertiary stress' consists of prominence principally produced by length and/or loudness. 'Secondary stress' involves an additional subsidiary pitch prominence. 'Primary stress' involving stressing of the most prominent of the most possible prominent syllable includes a principal pitch prominence. Thus, in Cruttenden's account, stress / accent are understood to correlate with different degrees of effort. This effort is manifested in the air pressure generated in the lungs (as a basis for the vocal-fold vibrations) for producing the tertiary stressed syllable and in the articulatory movements of the vocal tract for the primary stressed / accented syllable, as presented in Table 1. These production effects of stress result in various audible differences: a stressed syllable that is realized with pitch prominence stands out from its context (syllables that are unstressed). Thus, a high stressed syllable appears even higher if its neighbours are unstressed or low in pitch (known as 'emphasis for contrast', see Thorsen, 1979a). Another effect of prominence is that stressed syllables tend to be longer and louder than unstressed syllables, though experiments (e.g. Fry 1955, 1958; Isačenko & Schädlich, 1966) have shown that differences in loudness alone are not very noticeable to most listeners.

Later, Kohler (1977) and Beckman (1986) argued that for German and English the acoustic correlate of accentuation is not only intensified stress but a complex mixture of F0 variation, increased duration of syllables and words as well as increased intensity, due to higher subglottal pressure. Sluijter (1995) makes a starker distinction between stress and accent. In his terms, stress is a structural linguistic property of a word that specifies which syllable in the word is the strongest. Accent on the other hand is used to focus and is thus determined by the communicative intentions of the speaker. Thus, whereas stress occurs according to phonological word-rules, accent is manifested in the informational structure that a speaker wants to communicate.

To summarize, prosody enables one to highlight both at word level, meaning stress or lexical stress, but also at the level of utterance level, meaning accentuation. Compared to an unstressed syllable, a stressed one is louder, longer and more strongly articulated. A stressed syllable with an additional tonal movement has to be considered as pitch accent or, if it is the last pitch accent of an Intonation Phrase, as the nuclear pitch accent. In this thesis, I will use the term 'stress' to mean lexical stress and 'accentuation' (including accent and pitch accent as synonyms) to mean prominence at utterance level.

² Note that this determination is not based on linguistic categories e.g., noun or verb. Rather, the fact that e.g. pronouns are unlikely to receive stress is due to the fact that they often describe a referent that is already known by the interlocutor of a conversation. On the other hand, nouns often refer to elements that are new or somewhat important (cf. Chapter 2.3.2.)

Table 2, adopted from Baumann (2006:11) summarizes this and presents how different degrees of prominence are used in this thesis:

Table 2: description of the phonetic correlates of stress and accent used in this thesis, adopted from Baumann (2006:11)

No stress/accent	
Stress	syllable is louder, longer and more strongly articulated than an unaccented syllable
Pitch accent	additional tonal movement on or in the direct vicinity of a stressed syllable
Nuclear pitch accent	last pitch accent in an intonation unit

As we have seen, prominence at word level (stress) and utterance level (accent) have their correlates in language dependent phonological rules or in the intentional aspects of communication. In the following section, I will give an overview of the phonological aspects of intonation as well as systems which make it possible to describe the intonational contour within spoken language. Additionally, I will describe the functions of accentuation, based on both affect and intention.

2.3. The phonological aspects of intonation

2.3.1. Forms of intonation

Autosegmental and Metrical Phonology

In the literature, intonation has traditionally been described as either contours (giving the direction of the intonational pattern) or in terms of discrete levels (describing the degree of prominence of syllables). This has made it possible to carefully describe the range of an individual spoken intonational contour. One of these models, which will be used in this thesis, describes the

intonational contour according to the Autosegmental - Metrical (henceforth AM) theory of intonation.

Within this overall theory, "metrical phonology" is concerned with the organization of segments into groups of relative prominence. The theory describes the different prominence values and their relations within and between prosodic domains of different sizes (as e.g. intonation phrases, phonological phrases, prosodic words, feet and syllables) and the rhythmic structures of utterances (see e.g. Liberman, 1975; Liberman & Prince, 1977; Selkirk, 1984; Hayes, 1982; Uhmman, 1991 for detailed description of prominence relations). However, because the focus of this thesis is not children's acquisition of prominence relations, namely metrical aspects of prosodic prominence within different prosodic domains, I will focus on the principles of "Autosegmental Phonology", the second central part within the AM theory of intonation.

Autosegmental Phonology (e.g. Liberman, 1975; Bruce, 1977; Pierrehumbert, 1980; Pierrehumbert & Hirschberg, 1986) offers an abstract description for English intonation that allows the characterizing of all potential intonational patterns within this language. One important step in order to develop such a model was a careful investigation of the rules by which phonological representations are mapped onto phonetic representations (see e.g. Pierrehumbert, 1980). By doing this, not only a descriptive element for intonation was created, it was also possible to overcome the inadequacies of earlier description-models of intonational information. Thitherto, *The Sound Pattern of English* (SPE) by Chomsky & Halle (1968) (cf. Chapter 3.1.) was the standard theory of phonological representation in Generative Grammar³. In this work, Chomsky and Halle view of phonology was separate from other components of grammar. Instead, the underlying phonemic sequence of each sequence was transformed according to rules, its output was produced in terms of the phonetic form that is uttered by a speaker. However, the theory fits with the rest of Chomsky's theories of language in that sense that it adds a theory of phonology to his previous work on syntax. Thus, words are regarded as being split up in linear sequences of sound segments. These segments were represented in the form of unordered bundles of binary distinctive features, not only containing the 'segmental', but also the 'suprasegmental' information such as features for tone and stress. According to this, the SPE-model assumed prominence on individual segments. But, stress and accent are features that are not anchored in only one sound segment within a word but rather in the syllable. Additionally, the SPE – model only used binary features (like [+ stressed] or [– stressed]), which cannot be used to explain a relative and gradual concept like stress or prominence in general. Rather, these features are linked to syllables (at least in languages like German and English). And, as Pierrehumbert (1980) pointed out, whereas it is

³ It has to be noted that the AM-model is also a generative model in the sense that it is based on a limited number of features with which an unlimited number of tonal patterns can be built. However, this model does not assume that this is derived by innate mechanisms or rules.





possible to describe the articulatory realization of a sound with binary features, the linear arrangement of the SPE-model makes it impossible to represent a tonal movement within a single segment, e.g. a fall in pitch from high to low on a short vowel (e.g. [a]). What this means is that, although it is possible that two mutually exclusive features are realized within the same sound, this is not possible in the SPE-model, since a sequence of two features is not allowed within the same segment.

In the AM – theory of intonation, this problem was solved by separating the segmental and suprasegmental level. Instead, the two features, are organized on different `tiers`, i.e. the text and the tone tier. Although these two different levels are synchronized in that sense that they are reliant on each other, they can act autonomously as independent segments or ‘autosegments’ (‘Autosegmental Phonology’, see Goldsmith, 1976). Thus, the different features are independent of the syllable structure (and thus also independent of the syntactic structure).

An additional advantage of the system was the possibility to describe the intonation of spoken language. In this sense, intonational contours are described as sequences of high (H) or low (L) targets. These targets are allocated to the prominent elements of a word and are referred to as a ‘pitch accent’. Pitch accents are marked with a star ‘*’ following the tone, e.g. ‘H*’ for a high pitch accent. In cases in which the direction of an intonational contour is described (and thus, the accent consist of more than one tone), the two tones are combined by using a ‘+’ sign, e.g. ‘L*+H’ (indicating that the low tonal target corresponds with the lexically stressed syllable). Boundary tones, marked by a ‘%’, characterize the intonational contour from the last (nuclear) pitch accent to the boundary of the intonational phrase⁴. The following table summarizes this.

⁴ The number of syllables between the nuclear pitch accent and the end of a phrase can vary. Thus, it can happen that both the last pitch accent and the boundary tone occur on the same syllable. In this case, the annotations are summarized e.g., ‘H*%’.

Table 3: schematic representation of an utterance containing a rising-falling intonational pitch pattern within the utterance „good morning“ and a falling-rising intonational pattern on the utterance “on Tuesday” (partly adapted from Grice 2006). The first two rows indicate the F0-pattern on the corresponding utterance. In the third row, the stressed syllable is marked in capital letters. The fourth row shows the syllable structure representing the stressed syllable in the black area. The fifth row represents Autosegmental annotations of the pitch accent and the boundary tone .

H L	L H	tone
		F0-contour
goodMORning	onTUESday	text
		syllable structure
H* L%	L* H%	annotation

Beside an annotation of just high and low tones, it is possible to modify these two tones using operators in the form of ‘downstep’ and ‘upstep.’ If a high tone is considerable lower than the preceding high tone (but not as low as a L-tone), it is considered to be downstepped and marked with an exclamation mark before the downstepped tone, e.g., ‘!H*’. This feature often appears for example in listings⁵, described in (1)⁶:

⁵ This effect is sometimes also referred to as ‘declination’. Declination is typically assumed to be a phonetically effect, due to the decreasing amount of air in the lungs during the realization of an utterance. However, Pierrehumbert proposed that the phonetic declination effect exists, but also argued that the major contribution to the downdrift of utterances was ‘downstep’. In her view, this is a phonological effect and therefore under the speakers control (see Taylor (1992) for an overview)

⁶ If not otherwise stated throughout this thesis, capital letters indicate pitch accents. Since accents apply to syllables, not to words, we only capitalise the respective syllable.

(1)

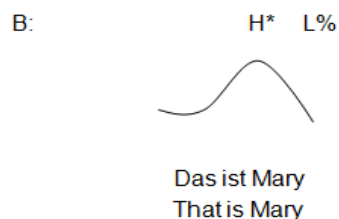
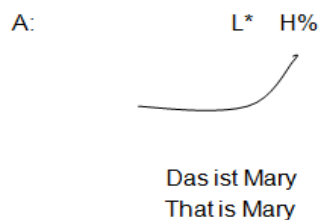
Zum Frühstück habe ich BROT, Marmelade und Bananen gegessen.
For breakfast, I ate bread, marmalade and bananas.

H^* $!H^*$ $!H^*$

An upstepped tone, indicated by a \wedge (e.g. $\wedge H^*$), indicates a tone that is considered as higher than the preceding tone. Overall, it should be pointed out that within the AM model, the order and thus the prominence of different pitch accents cannot be distinguished. For example, the nuclear pitch accent is simply described as the last fully-fledged pitch accent in a phrase; pitch accents before this nuclear pitch accent are described as 'prenuclear'. But both kinds of pitch accents are described in the same way within the model. Practically, the nuclear pitch accent tends to be the most important accent in the phrase, often signalling the main focus of the sentence. For example, in (1) above, the tone on "bread" and "marmalade" is described as prenuclear and the tone on "bananas" is considered to be the nuclear pitch accent – even in cases in which it does not carry the highest tone in the intonation phrase.

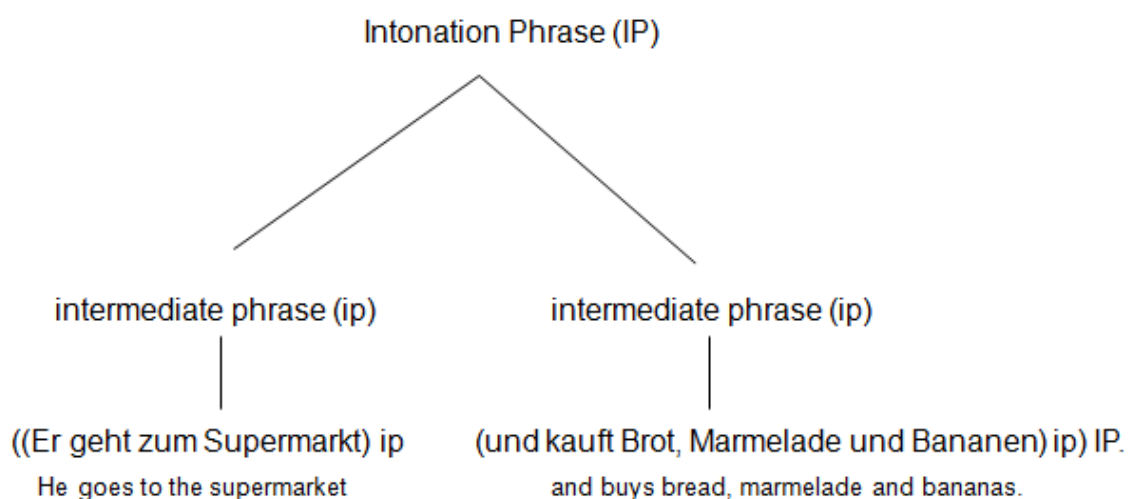
In the AM-model, it is possible to describe the way in which the two utterances differ in their intonational realization. Consider our example "That is Mary" from section 2.1., repeated in (2). Example A represents the intonational contour of that utterance with a rise at the end of the utterance, indicating disbeliefs about whether the person really is Mary. (2) B represents the pattern of a falling speech melody after a H^* - pitch accent in order to make a statement and introduce Mary to another person.

(2)



The AM-model makes it not only possible to describe the intonational pattern with which an utterance is realized but also the form of the utterance, that is the division of an utterance in several parts or 'phrases'. To do so, the model utilizes a third kind of tone – the phrase accent, described as ' - '. The phrase accent is always monotonal e.g., 'L-' or 'H-'. The phrase accent separates smaller units of intonation, also called 'intermediate phrase' (ip), which together form a part of larger 'intonation phrase' (IP). Intermediate phrases consist of one or more pitch accents plus a simple high or low tone that marks the end of that intermediate phrase. Thus, the phrase accent controls the F0 – movement between the last pitch accent of the ip and the beginning of the next ip. An utterance is allegedly built out of (at least) one Intonation Phrase, which consist of (at least) one intermediate phrase (see (3) based on Beckmann & Pierrehumbert, 1986).

(3)



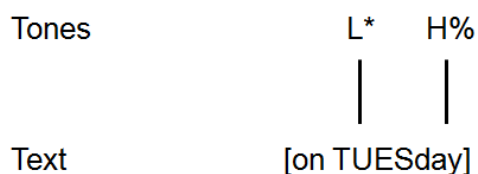
However, intonation and prosodic organization differ from language to language. The ToBI-system (Tones and Break Indices) was devised in order to develop a descriptive framework where it would be possible to describe the intonational pattern and the prosodic structure of different languages. ToBI is grounded in careful research into the intonation system and the relationship between intonation and the prosodic structures of the language examined. ToBI-systems have been developed for a variety of languages (e.g. for American English: MAE-ToBI – Mainstream American English; X-JToBI for Japanese or ToDI for Dutch). Each system is specific to a language variety and was developed by the community of researchers working on that language. The German variant (GToBI) was developed between 1995 and 1996 by researchers

from Saarbrücken, Stuttgart, Munich and Braunschweig (see Grice & Baumann 2002, Grice, Baumann & Benz Müller 2005 for an overview). Because this thesis is about German children's use and understanding of intonational patterns, German ToBI (G-ToBI) will be introduced in the following section.

GToBI

A (G)ToBI record works on at least three different levels of description. These levels contain labels for text, tones, and break indices. For the investigation about the role of intonation in language acquisition and its description, covered by this thesis, only information provided by the text and tones are important and will be focused on in the following sections. The association of the autosegmental tone and text tiers from Table 3 is given in (4).

(4)






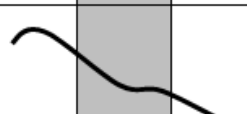


The text level gives information about the orthographic transcription of the spoken words. The tone level shows the perceived pitch contour in terms of tonal events such as pitch accents and boundary tones, and the break index level marks the perceived strength of phrase boundaries.

As mentioned in the previous section, pitch accents are associated with lexically stressed syllables. They are described as a starred tone placed within the limits of the accented word. They generally occur at local F0 minima and maxima. Table 4 summarizes and depicts the pitch contour of all possible pitch accent variations for the standard German variety⁷.

⁷ For transcription details see Grice & Baumann, 2002; Grice et al., 2005; and the GToBI webpage (<http://www.uni-koeln.de/phil-fak/phonetik/gtobi/index.html>).

Table 4: Schematic representation of possible pitch accents in German according to the GToBI system. The first column represents the syllable structure (the grey area indicates the stressed syllable) and the intonational contour. The second column describes the according GToBI annotation. The characteristics of the signal, both in terms of production and perception, are described in Column 3.

Contour	GToBI – annotation	characteristics
	H*	- peak accent - perceived as high
	L+H*	- low target before or at the onset of the accented syllable - peak often late in the accented syllable - sharp rise - perceived as high
	L*	- low target on the accented syllable - perceived as low
	L*+H	- low target on the accented syllable (following a high target on the preceding syllable) - perceived as low
	H+L*	- high peak followed by a sharp fall before the accented syllable - accented syllable perceived as strong pitch fall from high to low
	H+!H*	- the accented syllable is around the middle of the speaker's pitch range - preceding syllable is higher

Measuring and annotating intonational contours requires long-term training. Additionally, it is relatively time-consuming which is why studies in this area often contain small data-sets. Importantly, a transcriber has to set up rules that he follows throughout the annotation.

Grice et al. (1996) examined the overall inter-transcriber-consistency of a given data-set. In their study, 13 transcribers with differing levels of expertise labelled a diverse set of speech data using GToBI, labelling both pitch accents and edge tones. Their results suggest that, with sufficient training, labellers can in fact acquire sufficient skill with GToBI for large-scale database labelling. However, they found that there are in fact some confusing intonational contours, namely H* / L+H*. The disagreement between raters was mainly based on the relatively late peak in L+H*. Similarly, the contours L* / L*+H, L+H* / -L*+H and H* / H + !H* resulted in rater-inconsistency because of their similar pattern. However, although these contours cause some interdependent reliability problems, there is an indication that improved training might reduce the number of disagreements, since the developers were more consistent among themselves

than other labellers. The differences between raters were quite small indicating that non-experts can also gain operational skill with GToBI. The results from this study suggest that mechanisms that are quick to learn, provided by the system, is a necessary prerequisite for a system which is to be used for multi-site large-scale database annotation.

This subsection has provided an overview of intonation and a system to describe it, however, intonation of course also serves critical functions within spoken language. Children have to learn which form of intonation conveys which function, both in comprehension and production. What function prosody, and intonation in particular, fulfils with its different forms will be described in the next section. I will now discuss both paralinguistic functions, mainly provided by the physiological and physical properties that produce the speech signal, as well as linguistic functions of intonation.

2.3.2. Functions of intonation

Affective functions

In 1977, Morton observed remarkable similarities in the acoustic properties of the sounds used in competitive encounters. He found that the body-size of a species, conveyed by visual properties like erected hair, ears or tails can be directly associated with the pitch of the voice. There is a direct correlation between body size and the vibration rate of the vocal folds in mammals (i.e. the larger the body, the larger and heavier the vocal folds, the lower the pitch). Practically, to give the impression of being strong and dangerous, animals produce low-pitched sounds. On the other hand, to give the impression of being small and frightened, animals produce higher-pitched sounds⁸. Ohala referred to this association of the acoustic properties of vocalization and the intent of the vocalizer as “an inherent part of the human vocalization system” (Ohala 1983:13) and called this the ‘Frequency Code’. Later, Gussenhoven (2002) adopted Ohala’s term in order to explain the functions of intonation. In his view, there are two components: the phonetic implementation and the intonational grammar. The former is widely used for the expression of universal meanings that derive from three different ‘biological codes’, which he claims to be universal among languages. These codes derive from biologically determined conditions and explain what is universal about the interpretation of pitch variation. He defined the three codes as follows:

Frequency code: The term is an expansion of Ohala’s analysis regarding the widespread similarities in patterns of avian and mammalian vocalization in

⁸ Please note that, related to Morton, this also mimics infant vocalization. In an evolutionary sense, this is seen as being due to aggression reduction (see also Ohala 1980).

face-to-face competitive encounters. The frequency code explains universal gender specific differences in the sense that larynxes that have smaller size automatically contain smaller and lighter vocal cords. The result of this is faster vibration and higher fundamental frequency. The relation between larynx size and rate of vocal cords is typically supposed to be responsible for power relations. For example, vocalizations by dominant or aggressive individuals are typically low-pitched, while those by sub-ordinate or obedient individuals are high-pitched. A wide-spread explanation for this correlation is that lower pitch suggests that the speech organs are larger. However, higher pitch is commonly seen as friendly and polite (see also Chapter 9 for the role of pitch in child-directed speech). Within these categories, Biemanns (2000) found correlations between artificially produced speech, imposing either a masculine or a feminine voice. In this study, participants judged positive characteristics like being polite, non-aggressive and friendly on the 'feminity scale', whereas negative connotations of voice were judged more frequently as being on the masculinity scale.

Effort Code: The amount of energy that is needed for speech production can be varied in the sense that more effort will lead to more precise articulatory movements as well as more canonical and more numerous pitch movements (de Jong 1995). Excitement towards a certain event results in more sub-glottal pressure which then results in higher pitch movements. The speaker can use this in order to mark certain words or phrases in an utterance as 'special' or important. Additionally, another informational interpretation of the Effort Code is that of 'emphasis'. Speech directed towards children, in almost all languages, is produced with a wide excursion of pitch movements (see Chapter 9), which is often interpreted as the expression of 'helping'.

Production Code: This code associates high pitch with the beginning of utterances and low pitch with the ends. This originates from a correlation between utterances and breath groups. The subglottal pressure decreases throughout a breath group as the air is gradually used up. A new intake of breath means that the subglottal pressure becomes high again. Implications from this code are that high beginnings typically signal new topics whereas low beginnings continue a topic. Similarly, this holds for utterance ends: high endings signal continuation whereas low endings signal finality and the end of turn. Figure 1 summarizes the three codes.

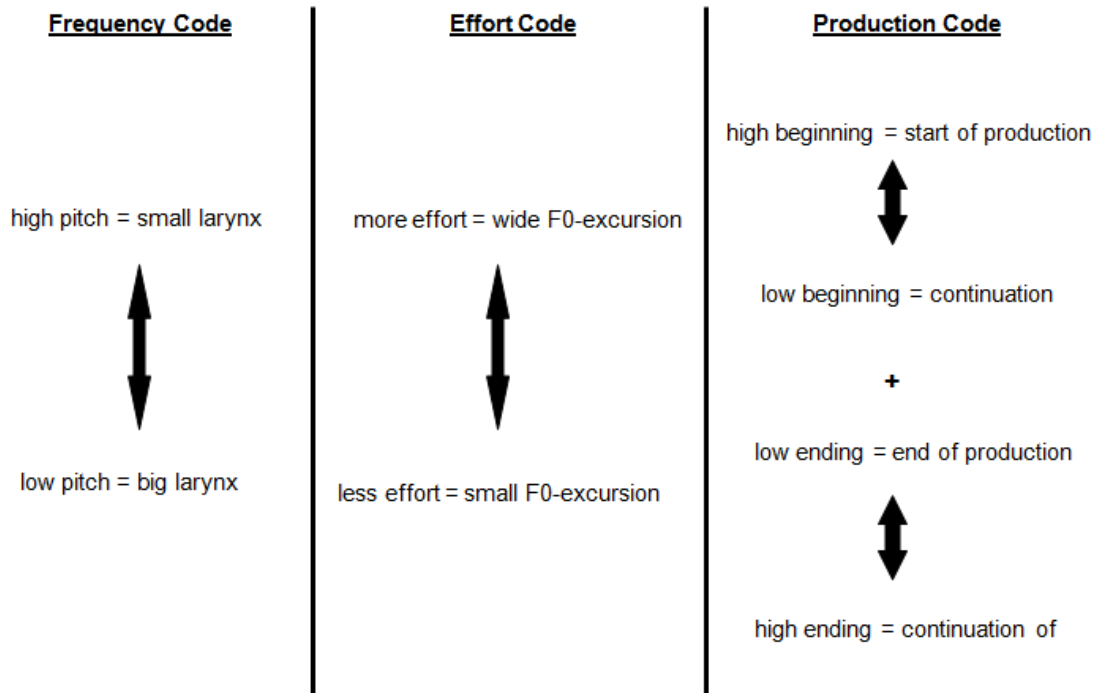


Figure 1: Summary of the biological codes.

According to Gussenhoven, biological codes are based on the effects of physiological properties of the production process on the signal. They represent aspects of the speech production mechanisms that affect the rate of vocal cord vibration. But, communication does not require that the physiological conditions are created. Rather, “it is enough to create the effects” (Gussenhoven, 2002:48). What this means is that the effects are not automatic, but have been brought under control. For example, by using the Production Code, Gussenhoven argues that a speaker does not need to think about an extra-exhalation phase in order to start a new topic. He only needs to raise the pitch of the first one or two syllables. However, whereas these implications, derived from the three biological codes are said to be universal to all languages, each of them also has implications for the grammar of intonation. These are supposed to be language specific. But, the two implications go hand in hand in the sense that linguistic meaning is potentially arbitrary, “although the form-function relations between tone and meaning frequently mimic the paralinguistic form-function relation employed in phonetic implementation” (Gussenhoven 2002:47).

What this shows is that prosodic cues like intonation can be realized “unconsciously” in order to express, for example, fear or happiness, due to the physical and physiological properties of the speech organs as proposed by Gussenhoven’s biological codes. In addition, Ohala (1983) noted that, for example, the frequency code can explain a number of cross-linguistic patterns in the use of pitch. For example, a high and/or rising pitch is used to signal yes-no questions because one is dependent on the other’s good will for the requested

information and the questioner is required to make some effort. When making a statement, one is certain about the situation that is being communicated and it does not require a significant amount of effort – which results in a low or falling pitch. This could lead to the conclusion that paralinguistic intonational meaning is completely universal, but there are indications that this is in fact not the case. For example, research on the vocal expression of emotion and the recognition of emotion (e.g. van Bezooijen, 1984; Scherer, 2003) has shown that although universal vocal cues for emotion exist, there are culturally specific variations. And, according to the linguistic means of intonation, listeners differed in their sensitivity to cues according to the frequency code, regardless of whether or not an utterance is a question. What this shows is that although biologically universal cues exist, which are responsible for a number of universal meanings (e.g. fear, happiness, and dominance), there are also other linguistic markings by intonation which happen intentionally. These cues belong to what Gussenhoven calls the grammar of intonation.

Intentional functions

As already mentioned, the distinction between the affective and the intentional functions is not easy. Speakers control the phonetic implementation of linguistic expression for a variety of reasons. For example, the effort code allows that, for special information, larger amounts of energy can be put into the realization of that information. In fact, a speaker does use these physical and physiological properties in order to lend meaning to utterances. Apart from the diverse linguistic and paralinguistic functions of intonation at utterance level, starting with the marking of sentence modality to the expression of emotional and attitudinal nuances, some languages like Chinese and Yucatec Maya use pitch variation and tonal contrasts for lexical and morphological marking in order to make distinctions at word level. For example, a widely cited example is the syllable 'ma' which has several meanings (*mother, hemp, horse, scold* as well as the expression of an interrogative particle). The exact meaning of this syllable is provided by its intonational realization. Additionally, in Bini, a language from the Niger Congo in West Africa, intonation is used as a grammatical marker: a change of tone marks the difference between tenses, e.g. low tone marking present tense and high or high-low tones marking past tense (see Crystal 1987: 172). By comparison, for intonation languages like English and German, pitch is not responsible to make morphological or lexical distinctions. Instead, pitch is only relevant at utterance level. Here, the syntactic structure and the intonational pattern are related to each other, though they do not correspond in a one-to-one mapping. For example, highlighting certain words or phrases or placing a prosodic break between two constituents can be used in order to disambiguate between different syntactic structures and are often the only ways to disambiguate them. Consider for example an utterance like 'The policeman followed the robber with the car'. In this statement it is unclear whether the policeman is sitting in the car using it to follow the robber or whether the robber is

using the car in order to escape from the policeman. When resolving such syntactic ambiguities, it has often been demonstrated that listeners are sensitive to prosodic features, especially intonation (see Warren et al., 2000). In this example, a break after the verb would indicate that the robber has the car whereas a prosodic break after the second NP would indicate that it is the policeman who is using the vehicle.⁹ Albritton et al. (1996) have argued that a speaker's awareness of ambiguity is the primary factor that influences the salience of prosodic contrasts in that speaker's production of ambiguous sentences. What this means is that both the knowledge of a speaker and a hearer are important in order to (a) understand that an utterance can be syntactically ambiguous, (b) to realize the utterance in a way that it can be perceived unambiguously and (c) to understand which information a Listener needs to make this utterance unambiguous.

What this shows is that intonation serves a very important function with respect to the informational structure of an utterance. Utterances can be divided into a more and a less informative parts. These "parts" have been named for example "given" and new information", "background and focus" or "topic and focus". Gundel and Fretheim (2004) pointed out that two different phenomena, namely, referential givenness / newness and relational givenness/ newness need to be distinguished. Intonation plays a role in marking both kinds of information structures. The first category deals with the pragmatic function of the intonational realization of referential expressions in an utterance. Specifically, referents can either function as background or focus. Their function is based on the structure of the existing discourse and the intention of a speaker. Whereas the more informative part of an utterance is linked to intonational prominence, the part that provides less informative, given, or background information is usually linguistically and intonationally less salient. Background information may originate from questions, with the answer to the question providing new information. Consider the following example

(A): What did you buy?

(B) [I bought]_{background} [bread]_{focus}

In this example, both "I" and "bought" in the answer are background information as they are already given in the opening question. The sought element in the question is the new information, the 'focus' in the answer, and thus that which is intonationally highlighted in speech (cf., Lambrecht, 1994). The relation of background vs. focus can be considered as largely equivalent to what is often referred to as new vs. given. That is, topical or background information is usually also given in the discourse, and focused information is also the new

⁹ Note that in this example, several prosodic cues have to be combined in order to resolve the ambiguity.

element in the discourse (for a detailed discussion of the differences see Gundel & Fretheim, 2004).

The givenness and newness of a referent in the discourse relates to its cognitive status in the mind of the listener (or the speaker's assumption about its cognitive state in the listener's mind). Depending on the degree of the assumed givenness / newness of a referent, speakers use different referential expressions. For nominal expressions, for example, this varies from using pronouns for referents in the current focus of attention to prosodically highlighted full noun phrases, (see Gundel, Hedberg, Zacharchki, 1993 for a detailed model). Furthermore, referential expressions for given and new referents differ in the extent to which they are prosodically highlighted. Referents can be either treated as given (see "I" and "buy" in the previous example) or new (as "bread"). In his model of Information Structure, Halliday (1967b) introduced the terms given and new treating them as a dichotomy: given information is presented by the speaker as being recoverable from the discourse context, new information is not. Chafe (1994:73) extends this binary distinction between given and new and defines three information states with respect to the activation cost a speaker has to invest in order to transfer an idea from a previous state into an active state. What he means is that a referent is given when it is already active in the listener's consciousness at the time of the utterance; if a referent becomes active from a previously semi-active state, it is considered to be accessible; if a referent is activated from a previously inactive state, it is new. Along these lines, Gussenhoven (1983) describes the meaning of nuclear tones in terms of information status as characterized with respect to a shared "background". He assigns accentuation as an indicator of the informational status of referents: a referent that is accented introduces new information into the discourse, whereas de-accenting is assumed to refer to already established or given referents.

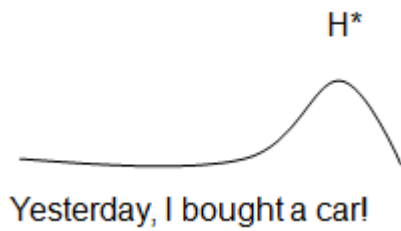
For West-Germanic languages like German and English, it is typically assumed that the placement of pitch accent is crucial for the marking of information status (Gussenhoven 2005). However, this distinction between accented and deaccented referents, conveying their status as either new or given, is a simple binary distinction. Several scholars have gone beyond this either-or distinction, whereby information is either given and thus deaccented, or new and thus accented. For example, Pierrehumbert & Hirschberg (1990) proposed that the distinction between given and new information is not dichotomous but rather that they are continuous and that different types of pitch accents convey information about which level of importance a speaker intends to assign to a certain referent. Pierrehumbert and Hirschberg pointed out that:

"a speaker chooses a particular tune to convey a particular relationship between an utterance, currently perceived beliefs of a hearer or hearers, and anticipated contributions of subsequent utterances. (Pierrehumbert & Hirschberg, 1990: 271)"

Thus, intonation is an important linguistic instrument that enables a speaker to structure his utterance taking into account what he thinks the listener does and does not know. In order to address the relevant information to a hearer, the speaker has to mark his utterance in an appropriate way. And, the hearer needs to have the ability to understand this marking. This involves not only knowledge about linguistic conventions, but also knowledge about the psychological status of referents within a conversation. Thus, in order to understand the communicative intentions of a speaker it is not only essential to know how to realize this information, but also to have a shared background, which is developing between the participants in a conversation throughout the discourse. Intonational features such as pitch accents, phrase accents and boundary tones can convey how a speaker intends a hearer to interpret the spoken intonational phrase with respect to: (1) what the hearer already believes to be mutually believed and understood (between the hearer and the speaker) and (2) what the speaker intends to make mutually believed as a result of subsequent utterances. Therefore, the kind of pitch accent provides information about the status of an individual discourse referent and its relationship to other referents specified by the pitch accents with which they are associated.

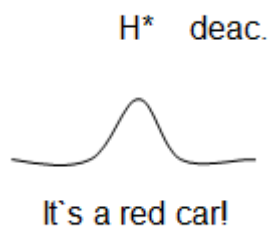
Whereas accenting or deaccenting a discourse referent appears to be associated with the speaker's desire to indicate the relative salience of accented items in the discourse, the type of pitch accent conveys other sorts of information status e.g., whether accented items belong to mutually held beliefs between the speaker and the hearer or whether they are inferable. For example, what a speaker says in the first sentence of a discourse may be considered to be completely new to the listener. This newness has to be marked in certain way. If the speaker refers to that matter again in one of the following sentences, the information has to be considered as given from the preceding discourse. The information has become part of the listener's knowledge. As a consequence, the speaker may use a different intonational contour when referring to that information a second (or third) time. To do so, all accent types can be used in order to transmit information from the speaker to the hearer about how the propositional content of the realized utterance is to be used. This is important in order to modify what the hearer believes to be mutually known between the two participants of the conversation. Pierrehumbert & Hirschberg summarize that "the meanings of the starred tones are shared among the different accent types" (1990: 301). In this sense, a H* - pitch accent is used to mark expressions that refer to elements in the discourse that are treated as new or (in Pierrehumbert & Hirschberg's terms) information that is to be added from the speaker's to the hearer's mutually held beliefs. Consider the following example.

(5)



After the referent “car” had been marked as new by the H* pitch accent, the corresponding referent is active in the discourse and can be treated as given in the realization of further expressions. Thus, it no longer needs to be accented (because both the speaker and the hearer know what is being talked about). Instead, the activated referent is deaccented, whereas other, newly introduced elements, get the H* pitch accent, as for example the colour of the car in the next example.

(6)



However, deaccentuation is only one appropriate marker for given or already established elements. Alternatively, Pierrehumbert & Hirschberg (1990) proposed that L* - pitch accents “marks items that S [the speaker] intends to be salient but not to form part of what S is predicating in the utterance”. For example, although “car” is already known by both the speaker and the hearer in a discourse-situation, the referent can be the most important part of an utterance. In order to mark this, the referent can be realized by a low pitch accent. This is shown in (7).

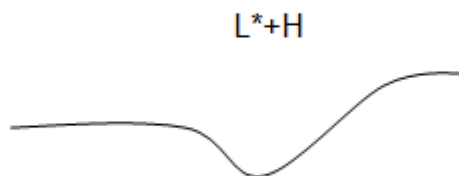
(7)



Furthermore, bitonal pitch accents are assumed to have a special pragmatic function. For example, all L+H accents “convey the salience of some scale [...] linking the accented item to other items salient in the hearer’s mutual beliefs” (1990: 294). In this sense, L^*+H accents are said to express uncertainty about a scale already evoked in the discourse. What this means is that this accent modifies or questions a common belief about a situation. Thus, it expresses for example uncertainty or incredulity, as in (8):

(8)

A: Alan's such a klutz.



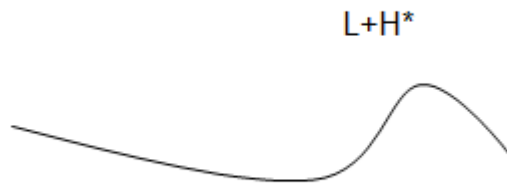
B: He's a good badminton player.

(taken from Pierrehumbert & Hirschberg, 1990:295)

Related to this, the $L+H^*$ pitch accents intend for the accented item to be mutually believed (in addition to mark correction or contrast). For example, in (9) the speaker assumes that the hearer has a certain piece of knowledge concerning the world (i.e. the weather in winter).

(9)

A: It's awfully warm for January!



B: It's even warm for December!

(taken from Pierrehumbert & Hirschberg, 1990:296)

For German, Baumann and Hadelich (2003) examined whether pitch accent type plays a role in the marking of different degrees of givenness (Chafe's levels of activation, e.g. Chafe 1994). Baumann and Hadelich presented adults with a variety of utterances containing target words that were marked with certain pitch accents. The words (or their referents) were either primed (auditory or visually) or were not primed. Participants were required to judge the appropriateness of the pitch accents placed on the target words. The results support Pierrehumbert & Hirschberg's (1990) analysis and show that H* was interpreted as the most appropriate marker for new information, while for given referents deaccentuation and L* - pitch accents were preferred. However, in this study no direct preference for certain pitch accents for accessible information was found - only one type of accessibility (situational accessibility) was tested. In a follow up study, Baumann and Grice (2006) used a similar procedure as in Baumann & Hadelich (2003) to investigate whether a certain pitch accent can be considered as appropriate not only for new and given elements (and thus already active in the listener's consciousness at the time of utterance, or inactive), but also for the appropriateness of a number of different kinds of accessible referents. To do so, they explored different relations between a textually given antecedent and any kind of expression that refers back (directly or via inference) to that given referent by using e.g. synonyms, hyperonyms or related referents within a scenario. They found that for information that can neither be treated as new nor given, but as something in between, H+L* pitch accents are considered as most appropriate.

Based on these findings, Baumann & Hadelich (2003) and Baumann & Grice (2006) presented the following mapping between the informational status of target referents and the appropriate intonational contour with which these target referents are realized

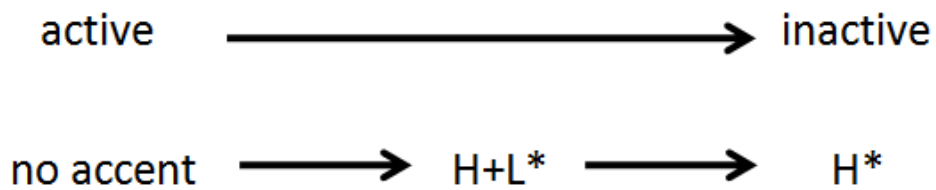


Figure 2: Baumann & Hadelichs (2003) scale of activation degrees (figure adopted from Baumann & Grice 2006:1655)

What these studies show is that both speaker and Listener in fact are sensitive to different degrees of the activation state of target referents. The intonational realization of target referents within a discourse is an essential instrument in order to convey the communicative intention of a speaker.

2.4. Summary

Intonation can fulfil very different functions within communication, ranging from marking information structure (semantic function) to conveying the paralinguistic properties of language, e.g. by communicating emotional states. Figure 3 summarizes this.

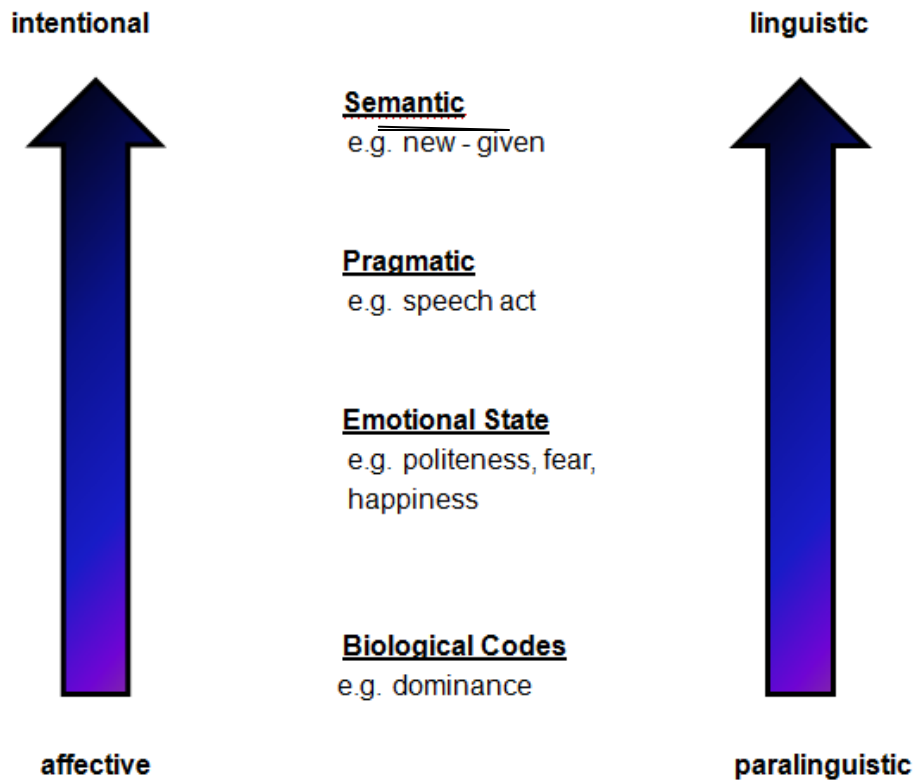


Figure 3: different functions intonation can fulfil (figure partly adopted from Grice, 2006)

At the level of semantics, intonation is often used to mark the informational structure within sentences. Thus, an utterance can be divided into an informative (containing new information) or less informative part (containing given information). We have seen that there is provision for a background-focus partitioning in which focus can be said to reflect an abstract notion of contrast between alternatives available in the discourse context (Rooth, 1992). The distinction between focus and background (or new and given information) can in many languages be marked by different pitch accents. For example, background is often marked by a lack of accent whereas focus is accented as there is always a major (nuclear) pitch accent within the focussed constituent.

At the pragmatic level, intonation is used to encode distinctions such as for example whether an utterance is intended as a request for information or as a request for the interlocutor to perform a particular action (e.g., Command). Four major categories of these communicative illocutionary acts has been defined: constatives, directives, commissives, and acknowledgments (Bach and Harnish, 1979; Searle, 1969); examples of which are statements, requests, promises, and apologies.

Intonation is also used to signal emotional states of varying degrees of intensity, affect and attitude. However, these emotional states are generally

considered to refer to function such as questions, statements and so on. Studies on their vocal realization have concentrated on non-discrete aspects of intonation, such as pitch range, rather than on phrasing and prominence relations or pitch accent type.

Although the expression of intonational meanings has been grammaticalized, it is claimed that there is a universal basis to this means of expression in the form of biological codes, the most established of which is the frequency code (Ohala, 1984), whereby high pitch corresponds to submissiveness or friendliness and low pitch to dominance or aggression. Two further biological codes, introduced by Gussenhoven (2002) are the effort code and the production code.

To summarize, intonation is active at many different levels of communication, in areas deemed purely linguistic as well as those considered more peripheral to linguistic inquiry. However, since the intonational expression of many functional levels occurs simultaneously, it is not possible to understand the expression of one level without taking into account the way the others are expressed. Thus, in the same way as a child has to learn the grammatical aspects of the morpho-syntactic level of a language, the child also has to learn the grammatical and the paralinguistic (in terms of both intentional and affective) aspects of intonation. The question arises at what point this process starts. As we have seen, it is not as easy to pull the accidental and the intentional aspects of language apart, as it appears that linguistic aspects derive from paralinguistic aspects. For example, are new elements marked by a high pitch accent because a speaker is excited about the new elements? Does this excitement result in a physiological reaction (i.e. deep breath, much air in the lungs) which then becomes conventionalized? Thus it seems plausible that a language learning child uses the paralinguistic properties of the intonational realization in order to understand the intention behind a certain behaviour. Later on, as language develops, the child can find patterns in this behaviour and eventually certain realizations are grammaticalized.

However, before I come to the empirical question of whether and in which way children use the intonational aspect of language in order to understand what another person is referring to and whether children can use intonation in order to learn language, I will give a brief overview about different approaches to the acquisition of language. Here, I will concentrate on the Nativist-Generative approach and the Usage-Based model.

3. Language Acquisition

3.1. The Nativist-Generative Approach

Interestingly, the Nativist-Generative approach emerged as a reaction to behaviouristic ideas. Here, Skinner (1957) presented in his famous book “Verbal Behavior” the idea that language acquisition could be explained with the same external processes that are used in order to explain behaviours in rats or pigeons. He claimed that these “methods can be extended to human behaviour without serious modifications” (Skinner 1957: 3). In his approach, language did not take into account any meanings, ideas or grammatical rules, i.e. anything that might be defined as a mental event. Instead, the methods that are used to control verbal behavior were based on classic conditioning. For example, let’s imagine a hungry pigeon in a box. The bird pecks on a button by chance – and receives food. After pecking the button several times, the pigeon will understand (or, in other terms, learn) that there is a connection between pushing the button and receiving food. What this means is that every time the pigeon pushes the button, it will receive positive reinforcement. According to this view, language learning is only one more type of conditioned learning by association. The first sounds an infant utters are strengthened by reinforcement, the mother reacts positively to that sound and the infant gets rewards. Thus, a verbal response is weakened or strengthened, depending on the type of consequences it may have: negative or positive. Both negative and positive reinforcement results in the full range of verbal sounds that are used in adult language. It was assumed that words and sentences can be learned in the same way. In this sense, sentences were just seen as a string of words without any structural relations between them. Thus, language is acquired by habit-formation via positive or negative reinforcement. In other words, a language-acquiring child can only rely on its environment in forms of positive or negative reinforcement. Thus, the study of language acquisition is reduced to the study of observables, i.e. to the observation of relations between input and output.

Overall, behaviourists treated physiological mechanisms (e.g. reflexes) and behaviour that is directly observable as a relationship between stimuli from the environment and the corresponding responses of the organisms. However, it is not clear exactly what happens between the occurrence of a stimulus and the immediate response. This process is considered to happen in a ‘black box’ in which nothing can be directly observed. Therefore, learning is defined without recourse to terms like ‘representation’ or ‘mind’, but simply as a relatively permanent change in a behavioural potentiality, a stimulus-response association resulting from temporal and spatial contiguity and/or positive and negative reinforcement of behaviour. Learning is viewed as a process of association and analogy formation that did not require any innate predispositions beyond a simple mechanism for forming associations and analogies in all domains of knowledge.

In 1959, Noam Chomsky argued in his critical review of Skinner's 'Verbal Behaviour' that the stimulus-response model is completely inadequate to explain the process of language acquisition. Chomsky offered several arguments: First, in order to understand the linguistic system in detail, it is necessary to understand what happens in the mind/brain of an individual speaker (which was considered as the 'Black Box' in Behaviourism). Only this can lead to an explanation of the most striking property of human language, the fact that we can generate infinitely many different expressions using a finite number of stored elements. In relation to this lack of clarity, Chomsky claimed that behaviourist explanations do not account for the production and comprehension of new sequences of words, which never receive any kind of positive reinforcement. Children (and adults) can also understand and utter sentences they have never heard before. As an example he offers the sentence 'Colorless green ideas sleep furiously'. Although the combination of these words is unlikely to have been heard before, and is not derivable from the input, it is possible to recognize this sentence as grammatical. This argument, dealing with the "Poverty of the Stimulus" (e.g. Chomsky 1980), claims that the grammatical competence displayed by children and adults cannot be simply derived from the input because the evidence in the language they hear around them cannot guide them to the abstract categories of language and its grammatical constructions.¹⁰ Nevertheless, as Chomsky pointed out, children learn fast and without any instructions on how to use language, without receiving any positive or negative feedback about their utterances with which to inform them about the grammaticality of their sentences¹¹. Based on this idea, he argued that the stimulus-response connection is not sufficient to deal with the problem of certain situations and the corresponding linguistic description. Instead, there must be some internal mechanism that allows the organism to choose new responses when facing certain situations. Chomsky's idea was that language can neither be described as a repertoire of responses nor can language acquisition be defined as the process of learning this repertoire. Instead, it is postulated that all languages share the same principles of grammar – the 'Universal Grammar' (UG). Internal mediating mechanisms facilitate language learning by setting

¹⁰ The argument about the "Poverty of Stimulus" is also known as "Plato's problem", which represents the question of how we account for our knowledge when environmental conditions seem to be an insufficient source of information. In Plato's "Meno" (470 BC-399 BC), Socrates tells Meno that there is no such thing as teaching. Instead, knowledge is a recollection of experiences from past lives. Socrates claims that he can demonstrate this by showing that even an uneducated slave boy knows geometric principles. Socrates states that he will teach the boy nothing, only ask him questions about the size and length of lines and squares, using visual diagrams to aid the boy in understanding the questions in order to assist the process of the so-called 're-collection'. The crucial point to this part of the dialogue is that, according to Socrates, although the boy has no training, he knows the correct answers to the questions – he intrinsically knows the Pythagorean proposition.

¹¹ This is known as the 'No negative evidence' argument (see e.g. Crain & Pietroski, 2001)

certain parameters¹². This parameter setting results in an activation of the specific properties of a language. This explains the fact that every sentence a person might understand or utter can be a novel combination of words. Additionally, children can acquire language rapidly, without any formal instruction, growing to correctly interpret constructions they have never heard before.

By introducing UG, the Nativist-Generative account draws a clear line between lexical items and syntactic rules that are applied to them in order to build sentences (e.g. Chomsky, 1993). Language is no longer interpreted as a system of habits, dispositions and abilities, rather it becomes a computational system based on rules and constraints that are specific to humans¹³. Such a view on language obviously led to a radically different interpretation of how knowledge of language is attained. As in all accounts of language acquisition, lexical items are arbitrary and thus have to be learned from the input. For example, children growing up in an English-speaking community need to learn that a four-legged, barking animal is called a 'dog', while children acquiring German need to learn that this animal is called a 'Hund'. There are no systematic relations between 'dog' or 'Hund' and the four-legged, barking animal. Thus, the lexical referents for objects or actions have to be learned from the input. The next step is to combine these language specific lexical items to sentences; that is to comprehend and produce sentences. To do so, several syntactic rules are needed. In the Nativist-Generative approach, these rules are assumed to operate within linguistic categories (e.g., noun, subject, object), that are said to be universal and supposedly the same in every language, rather than on concrete lexical items (e.g., 'dog'), that differ across languages. In order to acquire these (language universal) linguistic categories, the (language specific) lexical items need to be categorized. According to Pinker (1989), this is done using special linking rules which create systematic relations between lexical items and syntactic categories. For example, 'dog' refers to an animate thing and can thus be categorized as subject; 'tree' refers to an inanimate thing and can be categorized as object. Both the syntactic categories and the rules that link lexical items to these categories are said to be innate (Pinker, 1989).

However, the two principal arguments of this approach (learning lexical items from the input and the innateness of grammatical principles) are problematic. Children first need to categorize certain lexical items (e.g. 'dog') as predicates and heads, or nouns and direct objects, in order to activate the UG to

¹² Note that in the beginning, Chomsky claimed a special 'organ' of the brain that is supposed to function as a congenital device for language acquisition. This organ was called the 'Language Acquisition Device' (LAD). However, Chomsky has gradually abandoned the LAD in favor of the parameter-setting model of language acquisition.

¹³ In 2002, Hauser, Chomsky & Fitch claimed that that the sole quality of language that is unique to humans is recursion (defined as the capacity to generate an infinite range of expressions from a finite set of elements) (but see Gentner et al. (2006) for results on recursive understanding in European starlings)

set the parameter. But, how does a child know that what she hears being directed to her in a speech, qualifies for classification as a particular lexical item, such as 'noun' or 'verb'? "There is no direct relation between the types of information in the input and the types of information in the output: tokens of grammatical symbols are not perceptually marked as such in parental sentences or their contexts." (Pinker 1987:399). A potential solution to this problem is presented by the 'Principles and Parameters Account' (see Atkinson, 1992; Chomsky, 1999). In this account, the syntax of a language is described in accordance with general principles (the abstract rules or grammar) and specific parameters (i.e. markers, switches) that for particular languages are either turned on or off. For example, the head-direction parameter, i.e. the distinction between whether a language is head-initial (e.g. English: 'Mary has *seen* the book on the table') or head-final (e.g. German: 'Maria hat das Buch auf dem Tisch *gesehen*') is regarded as a parameter which is either on or off for particular languages (cf. next section). Thus, rules, as the properties of the specific language to which a child is exposed and pre-existing linguistic knowledge provided by the UG are supposed to link semantics and syntax.

"The suggestion is that children innately expect syntax and semantics to be correlated in certain ways in the speech they attend to, can derive the semantic representation by non-grammatical means (attending to the situation, making inferences from the meanings of individually acquired words), and can thereby do a preliminary syntactic analysis of the first parental utterance they process." (Pinker, 1989:360)

For example, in a sentence like 'The dog eats the apple', children are expected to categorize animated 'causal agents' like 'dog' as 'subjects' and inanimate 'affected patients' like 'apple' as 'objects'. They can then use this Subject-Verb-Object 'template' to produce and comprehend more sentences. The Nativist-Generative approach assumes that, due to the child's equipment with innate universal constraints on grammar, a child can find and match the language specific properties of universal categories with the specific settings in the domains of parametric variation. Since the input does not provide any perceptual markers of linguistic categories and rules, this matching cannot be achieved by purely perceptual mechanisms. In order to fill this gap, several bootstrapping mechanisms are assumed, defined as a link between input properties and knowledge of linguistic entities like 'noun' or 'subject of' provided by UG. This linkage itself is assumed to be part of an innate domain-specific inventory of capacities the child brings to the task of language learning. These bootstrapping mechanisms will be explained in more detail in the following section. Due to the topic of this thesis, my focus will be on the mechanisms of prosodic bootstrapping.

Bootstrapping mechanisms

The concept of bootstrapping underlies various proposals e.g., semantic bootstrapping (Pinker, 1987), syntactic bootstrapping (Gleitman, 1990) or the rhythmic activation principle for setting the head direction parameter (Nespor et al., 1996), as described above. The different kinds of bootstrapping mechanisms, characterized by the kind of information that serves as their input and the domain they help the child to break up, allows a language learning child to acquire several specific tasks in the language learning process. Although different linguistic fields are treated as unrelated and as having different responsibilities, all mechanisms have in common is that the child can, on the one hand, use cues from speech input or, on the other hand, use already established knowledge (and in turn use this for acquiring further linguistic knowledge - either within the same domain (autonomous bootstrapping; cf. Durieux and Gillis 2001) or within another domain (interdomain bootstrapping). For example, 'distributional bootstrapping' is assumed to compute non-prosodic segmented statistical properties of speech input at different levels of linguistic structure (phonemes, syllables, morphemes), in order to find syntactically relevant units in the input and assign these units to linguistic categories e.g., inflectional endings and function words typically belong to categories that occur frequently within languages. Additionally, due to their occurrence at the edges of words or syntactic phrases, they may provide information about clause-boundaries and information for the syntactic categorization of the elements with which they occur with (e.g., Gerken, 1996; Höhle et al., 2004; Maratsos & Chalkley, 1980; Mintz et al., 2002; Pelzer & Höhle, 2006).

'Semantic bootstrapping' as an association between semantics and syntax - as already mentioned above - addresses the question of how instantiations of linguistic categories and their relations are found. Semantic categories like 'action' or 'agent' are linked to syntactic categories like 'verb' or 'subject' which are part of the UG. Pinker (1984) assumes that children can construct a rudimentary semantic representation of input sentences with the help of context and their ability to understand the meaning of the words in those sentences. This allows them to identify basic semantic entities like 'agent' or 'action', etc. Accordingly, innate linking rules help them to connect the (newly acquired) semantic entities to the corresponding grammatical categories, which are said to be innate. And the specific morpho-syntactic features of the syntactic categories and relations in their target language can be identified.

'Syntactic bootstrapping' (e.g. Gleitman, 1990) allows the child to use the syntactic frames in which verbs, with their specific semantic component, appear. They then can use this syntactic frame to derive more (specific) syntactic functions of a specific word (or syntactic category). For example, a verb used in a transitive context has an agent and a patient and refers to a causative action, whereas a verb appearing in an intransitive context only requires an agent and refers to a non-causative action. Children can use this frame in order to learn the specific occurrence of a verb within its appropriate syntactic environment.

Gleitman and Wanner (1982) were among the first researchers to point out that prosodic information might help a child to discover the underlying grammatical organization of their native language. This assumption of the 'prosodic bootstrapping' approach, meaning that prosodic cues like stress, rhythm and intonation help the child segment the speech input into linguistically relevant units and categorize these units syntactically, underlies much work in acquisition research (for a review see Jusczyk 1997). It has been further proposed that prosodic information from the input can help identify word order regularities in the target language. For example, it is assumed that information about the rhythmic properties of the target language helps to set the correct head-direction parameter (Hirsch-Pasek & Golinkoff, 1996; Nespor et al., 1996; Guasti et al., 2001). To do so, the bootstrapping mechanism uses a correlation between the order of the head and its complement within a syntactic phrase and the position of the prosodic prominence within a phonological phrase. Typically, phonological phrases in head-initial languages assign stress to elements at the right edge of the phrase while phonological phrases of head-final languages have their most prosodically prominent element at the left edge of the phrase. This leads to different rhythmic patterns within the intonational phrase in these languages. Nespor and her colleagues proposed that children can make use of this correlation between stress assignment and head-setting parameter by way of an innate principle which they call the rhythmic activation principle (Nespor, Guasti & Christophe 1996).

Similarly, research in the area of prosodic bootstrapping follows on from the idea that prosodic information might help the child to identify units in the speech stream that correspond to syntactic or lexical units in the language. In many utterances, syntactic boundaries are marked by specific prosodic boundary markings e.g., lengthening of the final syllable, pitch movements and pausing at the boundary. Thus, it is suggested that infants are sensitive to these acoustic features that serve as boundary cues from an early age. Several studies have shown that infants around the age of 6 months react differently to speech strings with pauses inserted at syntactic clause or phrase boundaries than to speech strings with pauses inserted within clauses or phrases (Jusczyk et al., 1992; Hirsh-Pasek et al., 1987).

Directly associated with the segmentation of phrases using acoustic cues, it is assumed that children at this age start to segment their input into smaller units than clauses and phrases, namely, words. But, to do so, they need to glean some information about where a word starts and where it ends. This is complicated as in spoken language, assimilation and elision processes affect words. Additionally, in contrast to the cues which were discussed as being signals for clause and phrase boundaries, there are no clear acoustic-phonetic cues associated with word boundaries (e.g., Cutler 1994).

Bootstrapping accounts provide a natural explanation for areas of seemingly error-free acquisition. This holds especially for those accounts formulated within the framework of UG. If a parameter is set by the identification of specific input patterns, the corresponding linguistic knowledge is established as soon as the child has the perceptual capacities at her disposal and has

identified the necessary input features. This can happen long before the child is able to produce utterances that indicate that a specific grammatical property has been acquired, as shown for instance in the domain of the acquisition of word order regularities. Bootstrapping accounts postulate interfaces between different domains or modules of the language system or between subcomponents of a domain. These interfaces may be responsible for parallel acquisition in different domains of language.

As Höhle (2009) points out, the problem is the reliability of the individual acoustic cues that serve as boundary markers.

All these acoustic cues, taken alone, serve quite different functions within the linguistic system [...]. For example, F0-contours are associated with pragmatic functions like signalling whether an utterance is meant as a question or as an assertion. Lengthening is a relational property that can only be computed in comparison to the same syllable not produced phrase finally. The absolute duration of a single segment does not give any information concerning lengthening as segments differ with respect to their inherent duration, whether they appear in a stressed or an unstressed syllable and whether the language makes use of length as a phonologically distinctive feature. Pausing is not only related to boundaries but can also be an indication of some problem in the production process such as, for instance, problems in lexical access." (2009:373)

Furthermore, most bootstrapping mechanisms do not link units to one particular category. Rather, they are treated as an initial guess about the possible categories and units of the input. Due to the fact that units in different linguistic domains do not map onto each other in a one-to-one fashion but only show a more or less close correlation, the child has to overcome the application of a bootstrapping mechanism at some point during development. That is, for instance, if the child kept relying exclusively on a metrical word segmentation strategy, an English or German learning child would never come to a correct segmentation of iambic words or of typically unstressed function words. But there is evidence that by the end of their first year, children already treat iambic words as units (Jusczyk et al., 1999) and recognize high-frequency function words as units that are separable from their contexts (Höhle & Weissenborn, 2000; Höhle & Weissenborn, 2003). This suggests that children have integrated additional information into their segmentation routines, such as for instance allophonic information (Jusczyk et al., 1999), phonotactic information (Mattys & Jusczyk, 2001), and knowledge of frequently co-occurring patterns in the input (Saffran, Aslin & Newport, 1996). What this means is that children do not only just use one cue, but rather a mixture of cues in order to analyze the speech they hear (I will come back to this issue in Chapter 2.2.2).

However, as we have seen in Chapter 2.3.1., information that is provided by prosody does not reflect a one-to-one mapping between one special prosodic form and a corresponding special syntactic form. Instead, prosody as interplay of several physical and physiological properties provides information about different functions. Whereas the generative approach does not take into account this form-function mechanism, the Usage-Based approach of language acquisition, presented below, seems more suitable for integrating intonation as a cue that children use in order to understand and to learn language. This approach is based on the intentions a speaker wants to convey to a hearer. To do so, he organizes his utterances in the appropriate way. As we have seen in Chapter 2.3.2., intonation is an important instrument for organising the speech stream into more or less informative parts, but also in order to convey para-linguistic information. In the following section, the view of the Usage-Based approach will be described in more detail.

3.2. Usage-Based Perspective

Whereas the Nativist - Generative Approach assumes that innate linguistic categories process the linguistic input and that these categories (or principles) of core syntax do not have to be learned because they are there from the very beginning, some researchers argue that it is impossible to acquire language-specific properties by the activation of innate learning mechanisms. Instead, these features have to be learned and processed from the input over years or, in other words, language should be possible to learn by using language (e.g. Elman et al, 1996; Tomasello, 2003). Thus, the term “Usage-Based” was established by Langacker (1987) who assumed that the linguistic system of an individual speaker is established by the use of language, i.e. in concrete usage events or utterances. The linguistic system should be built-up from usage events of particular symbolic units. With increasing linguistic experience, more abstract linguistic patterns may evolve through using them. Thus, the Usage-Based approach can be directly applied to language acquisition (cf. Abbot-Smith & Tomasello, 2006; Tomasello, 2003). According to this approach, psychologists and linguists no longer think about the acquisition of language as isolated association-making and induction, but rather as a development in which the process of language acquisition is integrated and embedded in diverse cognitive and social-cognitive skills¹⁴. In this view, two sets of skills are of particular importance: intention reading and pattern-finding.

¹⁴ In discussing the emergence of language, Tomasello (2008) argues that human cooperative communication rests on a psychological infrastructure of ‘shared intentionality’ (joint attention, common ground)

Intention reading

A (communicative) intention can be defined as one person expressing a communicative device to another person in order to share attention with that person about some third entity (Tomasello, 1998a). In order to understand what a speaker is referring to with the help of linguistic symbols, it is of the utmost importance to know and to understand what that person has in mind when uttering that linguistic symbol or, in other words, to understand the person's intentions. Intention reading or, more importantly for language learning children, the understanding of other persons as intentionally acting agents (broadly defined as 'theory of mind') emerges around a child's first birthday (Tomasello, 1995a) and consists of various skills. It includes the idea that sound-making is not just about making noise, but that it has an underlying intention. Intention reading allows one a range of abilities: to share attention with another person towards objects and events of mutual interest (Bakeman & Adamson, 1984), to follow another's attention and gesturing to objects and events that are outside the immediate interaction (Corkum & Moore, 1995), the use of gestures in order to point, show or direct attention to objects (Bates, 1979) and, most important of all, the ability to imitate others' intentional actions but also to imitatively learn the intentional actions of others. For example, children between 9 & 12 months follow an adult's gaze and begin to look reliably to where an adult is looking (see Meltzoff & Brooks, 2007). The child comes to understand that an adult is not looking at an object for the sake of it, but that something about that object is interesting. Based on this newly detected potential for observation, infants start to observe that adults not only look at objects but also act on them. In a second step, they start to imitate this behavior and act on that same object in the same way as the adult. What makes this step so important in terms of intention reading is that this behaviour reflects a triadic relationship between the infant, the adult and the object. To achieve this, the child needs to coordinate her behaviour both towards the adult and the object. The infant now understands that others, as well as themselves, are intentional agents (Tomasello, 1995a).

For the use of intentionality within the process of language acquisition, three main stages of development are of particular importance. First, understanding others as intentional agents appears in an activity of 'joint attention'. Joint attention is generally known as the process by which one individual draws another individual's attention to a stimulus using non-verbal cues (e.g. gaze, pointing) as a signal. In order to achieve a goal e.g., to communicate with each other, the interlocutors have to be aware of the communicative content or discourse. For young children, this discourse can typically be an object that they act on. For example, imagine an infant and her mother playing with a ball on the floor (i.e. they are in a triadic situation). This situation could also be described as the joint attentional frame; the child understands (because of the newly acquired ability to see others as intentional agents) that her mother is attending to both her and the object. Interestingly, for the first time the child is situated in the same position as her mother: she is attending not only to the object, but also to her mother.

Importantly, the joint attentional frame gets its existence from the understanding that the observed object is part of the joint attentional frame. The sofa in the corner or the tree outside the window is not what the mother and the child are referring to in the here and now. This is not part of the joint attentional frame or the goal directed activity. In other joint attentional activities the object can of course change e.g., when watching a bird in the tree. This process of understanding what both you and I are attending to in a certain situation is the basis for the establishment of a common ground. And in turn, with the emergence of a common ground between two interlocutors, an individual can understand what another is referring to in a particular situation by using certain linguistic symbols. In other words, one can understand communicative intentions, the second important skill in order to read others' intentions. Within the joint attentional frame, a child understands that her mother is referring to the particular object that both individuals are concentrating on, in our example the ball. In the same way as the child understands that actions within the joint attentional frame are intentional to the object in this frame, the child also understands that communicative acts within the joint attentional frame are intentional to the object. For example, when the adult makes a sound, the child understands that this sound is not some kind of spontaneous and disconnected noise, but that it refers to the object on which both individuals are concentrating. Thus, sounds become language for young children when they understand that the adult is making that sound with an intention. In order to identify and to understand the referent of a linguistic symbol, it is necessary that the child can read the communicative intention, uttered within the joint attentional frame. This shows the importance of the joint attentional frame for learning communicative and linguistic intentions.

To summarize, at around 9-12 months of age, human infants begin to understand that other people act as intentional agents in order to achieve a goal. Additionally, having acquired this understanding, infants themselves become intentional agents. This enables them on the one hand to understand adults' intentional behaviour towards objects and activities within a joint attentional frame (and subsequently also toward objects and activities outside the joint attentional frame), and on the other hand to understand an adult's intentional state toward themselves and to their own intentional states. Finally, the infants themselves start to act as intentional agents toward objects and others.

Once the process of understanding others as intentional agents has started, this allows the child to use some new and species-unique forms of social learning. This tertiary stage within the use of intentionality for acquiring a language is also known as 'cultural learning'. The underlying learning-process is based on children's ability (both cognitive and physiological) to produce language on their own. Children do not only want to understand communicative intentions, they also want to realize them on their own in order to achieve a goal. In this sense, their understanding of the different processes involving the joint attentional frame and of communicative intentions makes a child more careful in observing other people when trying to achieve their goals. This leads to an imitation of individuals in the close environment in order to achieve goals of their own.

The main problem that the child is faced with in this situation is the problem of role reversal imitation. The learning- and imitation-process of intentional actions is relatively simple - the mother's and the child's treatment of an object occurs in parallel (the child sees the mother use her hands to lift up the ball, and therefore the child uses her hands to lift up the ball). The child can simply replace the adult with herself. But, communicative intentions are more complicated. When an adult confronts the child with a novel communicative symbol in order to refer to an object and the child wants to attend imitatively to that object, the situation changes.

“The reason is that in expressing communicative intentions in a linguistic symbol, the adult expresses her intentions towards the child's attentional states. Consequently, if the child simply substitutes herself for the adult she will end up directing the symbol to herself – which is not what is needed. To learn to use a communicative symbol in a conventionally appropriate manner, the child must engage in role reversal imitation: she must learn to use a symbol toward the adult in the same way the adult used it toward her.” (Tomasello, 2003:27)

What this means is that a child is faced with two different tasks. First, she has to learn to use a symbol for a certain object or for a certain situation, and second, she must use this symbol directed to the adult in the same way that the adult used it to her. Thus, she must replace the adult with herself as the target of an intentional, communicative act. Once this is done, the communicative symbol is understood inter-subjectively within a linguistic group. This also means that the linguistic symbol is shared between all members of that group. The Usage-Based approach treats this process as a social-pragmatic act (e.g. Tomasello 2003). The child comes to understand that using linguistic symbols is a social-act between two (or more) interlocutors, attending to an object together in a triadic way.

Pattern finding

According to usage-based linguistics, language structure can be learned from language use by means of powerful generalization abilities (e.g., Elman et al., 1996; Tomasello, 2003). This means that children do not only have to understand that linguistic symbols are part of a social-pragmatic act, in which the interlocutors interact with each other. In order to learn and to understand the grammatical dimensions of language, they need some additional prerequisite skills, namely 'pattern-finding skills' or 'categorization'. Recent evidence suggests that language learners can use statistical properties of linguistic input to discover structure, including sound patterns, words, and the beginnings of grammar. These abilities appear to be both powerful and constrained, such that

some statistical patterns are more readily detected and used than others. Several researchers have found that young children have excellent abilities at finding pattern in the auditory material that they are exposed to even before they start to speak. For example, Saffran, Aslin and Newport (1996) could show that 8-month-old infants could already segment words from fluent speech, based on the statistical relationships between neighboring speech sounds. The authors claimed that word segmentation is based on statistical learning. Although they concluded that infants have access to a powerful mechanism for the computation of statistical properties of the language input, these results can also be interpreted as indication for infant's prelinguistic ability to find patterns in auditory stimuli. Other studies showed similar effects with tri-syllabic words (e.g. Marcus et al., 1999) and with older children (e.g. Gomez & Gerken, 1999).¹⁵ Pattern finding seems to be necessary in order to understand linguistic mechanisms. The more often a lexical item is used in the input, the better the child understands its function. And, the better the function of a specific item is understood, the better the child can detect a pattern for that construction. For example, 'Where is the ball?' can be substituted into 'Where is Daddy?' or 'Where is the juice?' or simply 'Where is X?' This means that "fluency with a construction is a function of its token frequency in the child's experience" (Tomasello, 2000:453). The central cognitive phenomenon that is assumed to be responsible for the 'organization' of this experience is called 'entrenchment'. Frequently occurring repeated structures leave memory traces which are stabilized the more often this structure recurs. Entrenchment applies to both smaller units (e.g. morphemes, words) and "prepackaged" larger units or constructions. However, repetition on its own is not sufficient for understanding more general information. In order to generalize and form categories, the mind must recognize similarities as well as dissimilarities. It filters out aspects that do not recur, and registers commonalities by comparing stored with new units. New units are categorized along those dimensions wherever similarities with stored units are detected.

This result in children starting to communicate with so-called 'Holophrases':

"When they attempt to communicate with other people they attempt to produce (i.e., to reproduce) the entire utterance even though they often succeed in (re)producing only one linguistic element out of the adult's whole utterance. This kind of expression has often been called a 'holophrase' since it is a single linguistic symbol functioning as a whole utterance, for example, 'That!' meaning 'I want that' or 'Ball?' meaning 'Where's the ball?' (Tomasello 2000: 65).

¹⁵ These results were already discussed in Chapter 3.1. with alternative interpretations

Thus, the Usage-Based approach assumes that children, learning their first language, do not operate with adult-like categories, but rather with a psycholinguistic point of view. For example, when the child says 'Wanna play horsie', it is possible that she understands initial clauses in general (as assumed by the generative view). On the other hand, it could also be possible that the child just understands something like 'Wanna' + 'wanted action'. Thus, to resolve this issue, one has to look at the underlying linguistic representation. The Usage-Based-approach deals with the question of whether these representations consist primarily of concrete, item-based utterance schemas, or whether they are based on more abstract linguistic 'rules' (plus a lexicon to fill these with semantic content).¹⁶ Research done in this field suggests that most of young children's early language is not based on abstractions of any kind, but that children produce item-based structures with highly constrained 'slots' e.g., 'X VERB Y' (see Tomasello, 1992; Pine & Lieven, 1997; Lieven, Pine & Baldwin, 1997; Lieven, Behrens, Speares & Tomasello, 2003). As Tomasello (2000) argues, children's early multiword speech shows, a functional asymmetry between constituents, e.g. one word or phrase that seems to structure the utterance in the sense that it determines the speech act function of the utterance as a whole, with the other linguistic item(s) simply falling into variable slot(s). This kind of organization is responsible for what has been called the 'pivot look' of early child language, which is characteristic of the majority of children learning most of the languages in the world (Braine 1976; Brown 1973). Examples of early multi-word productions are: 'Where's the X?', 'I wanna X', 'More X', 'It's a X', 'I'm X-ing it', 'Put X here', 'Mommy's X-ing it', 'Let's X it', 'Throw X', 'X gone', 'I X-ed it', 'Sit on the X', 'Open X', 'X here', 'There's a X', 'X broken'. By generalizing this pattern, children's early grammars could be characterized as an inventory of utterance schemas that revolve around verbs, so called 'verb-island constructions'. Similar results have also been found for languages other than English (e.g. see Behrens, 2000 for Dutch; Allen, 1996 for Inuktitut; Gathercole et al., 1999 for Spanish; Stoll, 1998 for Russian; but see Lieven et al., 1997; and Akhtar & Tomasello, 1997 for frames based on pronouns).

Related to this, the question arises how children come to acquire more complex grammatical constructions. The answer lies in the nature of language according to the Usage-Based framework. Here, language is understood in terms of constructions. Like lexical items, syntactic constructions have a form and function. It is assumed that grammatical constructions are organized and represented in a network of related constructions (although it is stressed that constructions are not described as being derived from one another or from the same underlying construction). The basis for this assumption is that complex constructions derive from simpler ones. Due to the fact that phonemes and morphemes are also considered as constructions (Goldberg, 1995), an English

¹⁶ Utterances like 'wanna play horsie' are simply treated as adult-like utterances in the generative approach

plural –s and a noun are seen as combining to build a more complex construction (dog + s = dogs). Thus, learned words can already be put together in an indefinite number of constructions. For example, once a child has acquired the referents for dog and cat and learns under some circumstances that ‘The dog chases the cat’, this construction, categorized as ‘X verb Y’, can be used for other transitive constructions. Because the former construction inherits some general features from the latter (e.g., word order), the child uses this ‘template’ for other situations in which she wants to describe that ‘X verb Y’.

Opposed to generative grammar approaches, which claim that language acquisition is already complete by a very early stage, the Usage-Based approach assumes that the language acquisition process is continuous into adulthood. Adults and children at some point can form novel phrases because they have developed abstract constructions and they can use them to form new lexical items and rearranging familiar lexical items. Proponents of the Usage-Based view suggest that we arrive at this point by storing individual utterances as exemplars. Each utterance we hear is compared to the ones we have already stored (e.g. ‘X verb Y’). If the utterance we hear is identical to an existing exemplar, this exemplar’s representation will be strengthened. If it is not identical but is semantically and syntactically similar to an existing exemplar, it will be stored independently but close to the existing exemplar. Exemplars that are stored close to one another can then be compared, and, given sufficient commonalities, can be abstracted into a ‘schema’. The schema represents the parts that the individual exemplars have in common and is strengthened with each utterance that can be categorized and stored as an instance of it (cf. Bybee, 2006; Langacker, 2000). However, this has not been investigated further and it is not clear exactly on which grounds the processing system determines that individual exemplars are sufficiently similar to one another in order to be stored close by and to form an abstract schema. Future research will have to show how much this similarity is determined by factors such as meaning, form, or non-linguistic context (see e.g. Ibbotson & Tomasello, 2009).

To summarize, according to the Usage-Based approach, children in the early stages of language learning use language the way they have heard it used by adults around them. They acquire an inventory of item-based utterance schemas, with perhaps some slots within them built up through observed type variation in that utterance position. More abstract linguistic categories and schemas arise when children have achieved sufficient linguistic experience, in particular usage events to construct adult-like linguistic abstractions. It follows that the linguistic input plays a big role in linguistic development. The more frequently specific lexical items are used in an item-general, abstract pattern, the more lexically specific this pattern becomes. Lexically specific patterns or chunks can then gradually be turned into processing units that are independent of the abstract pattern (e.g. Bybee, 2006; Goldberg, 2006; Langacker, 1987).

One of the main problems in research on language acquisition is the logical question, how children can learn, produce and understand an unlimited number of sentences even though they hear only a finite number of sentences from their target language. Whereas generative linguists assume innate principle and

parameter settings, which constrain the space available to children for making hypotheses, Usage-Based linguistics focuses on the social-pragmatic and general cognitive skills of young children. These skills enable the child to understand the intentional mechanisms behind the use of language. The main difference between these two approaches is thus that the former one assumes an innate learning mechanism, based on a complex system of parameter settings and linking rules, whereas for the latter the acquisition of language is based on the use of language.

3.3. The role of intonation in the two approaches

The previous sections provide a brief overview of the different theories that have been devised to explain the language acquisition process. However, neither of these two models provides any specific information about the role prosody, and intonation in particular, plays in the process of language acquisition. The Nativist-Generative approach sees an influence of prosody only in order to help a child set certain parameters (cf. prosodic bootstrapping). For example, it is proposed that children can use the correlation between stress assignment and head-setting parameter by way of the rhythmic activation principle (Nespor, Guasti & Christophe, 1996). Additionally, in terms of marking the main prominence at the level of utterance (Focus-marking), Chomsky & Halle (1968) presented two rules: the 'Compound stress rule' and the 'Nuclear stress rule (NSR)'. The first rule proposes that stress is always assigned to the left-most stressable vowel in nouns, verbs, or adjectives, e.g. 'BLACKbird'. In a major constituent, e.g. 'the 'black BIRD' stress is assigned to the rightmost stressable vowel. The authors claimed that stress assignment is completely automatic once the syntactic structure is specified. Related to this, the NSR goes back to Newman (1946) who proposed that within an intonational unit, the last heavy stress is associated with the nuclear heavy stress. Based on this, Chomsky & Halle therefore formulate the NSR as a cyclic rule, that is, a rule that can be applied recursively.

“Once the speaker has selected a sentence with a particular syntactic structure and certain lexical items (...) the choice of stress contour is not a matter subject to further independent decision. (...) With marginal exceptions, the choice of these is completely determined as, for example, the degree of aspiration.” (Chomsky & Halle, 1968:25 f.)

However, as we already know from the previous chapters, prominence cannot be linked to the syntactic structure of an utterance. According to the Compound Stress Rule, stress may shift in certain constructions e.g. FIFteen vs fifTEEN girls. And, syntactic structures do not behave as predicted (consider our “This is Mary” example). Overall, the mechanisms regarding the intonational system mainly exist in order to understand the overall syntactic structure of a language, but not its variety of possible intonational contours.

Within the Usage-Based approach, the construction is one of the most important elements in order to acquire a language. Here, language is understood as constructions that have a form and function. When we have a closer look at construction in this sense we realize that the intonational form of an utterance is part of that construction.

“[...] there is one word or phrase that seems to structure the utterance in the sense that it determines the speech act function of the utterance as a whole (often with help from an intonational contour), with the other linguistic item(s) simply filling in variable slot(s).” (Tomasello 2000:66)

Remember our “Mary” example, here repeated as (10)

(10)

A: L* H%



Das ist Mary
That is Mary

B: H* L%



Das ist Mary
That is Mary

The lexical, and the resulting syntactic construction, are identical because both utterances consist of the same three words. What differentiates these two

constructions is their intonational realization. Thus, the intonational form takes over a function – and this function is dependent on its (intonational) form. What this means is that the pure formal treatment of intonation fits perfectly into the model of the Usage-Based approach. And, also in terms of the acquisition of language, this approach seems to be perfect for intonation. If we have a closer look at the tasks that intonation fulfils within the communication between two persons, as already described in Chapter 2.3.2., we can see that one principle task of intonation is to convey information about the cognitive status of a referent in the mind of the hearer and the listener. For example, if I would like to tell you that I bought a car (let's assume I never had a car before, we have not talked about a car or any other vehicle in our recent conversations and we are not surrounded by cars – simply put, I as the speaker assume that you do not have any picture of a car in your mind), I make the utterance: 'I bought a car!' In order to make sure that you really understand what I am talking about (and because this is what I want to tell you – i.e. it is my communicative goal), I have to make this part ('car') within my utterance especially salient. I do this by accenting it. From this moment on, the referent 'car' is activated in our discourse (or joint attentional frame) and I no longer have to accent it. Instead, any new element in the continuing conversation is accented (e.g. 'It's a BLUE car!').

What this means is that, in order to convey information in the best and most effective way, I have to know what you know, as well as what you know of what I know and so on. Thus, I have to make sure that you can read my communicative intentions. And of course, we both need the same background (or linguistic environment) in order to understand the communicative intentions, provided by intonation. I have to know what we are talking about and what the content of our joint attentional frame is. When I want to change this frame, I have to mark it in a special way. And, at some point, I must have learned this knowledge (we could also say these 'mind-reading abilities'). Within the two approaches to language acquisition, the generative approach seems inadequate for doing this. As mentioned before, prosody cannot be linked to single-segments but is a property of the situation and the social-pragmatic background of the speaker, the utterance and the context ('I bought a blue CAR' vs. 'I bought a BLUE car'). On the other hand, the Usage-Based account seems to be the perfect approach in order to understand the nature of intonational development. As we have seen, this approach assumes that children acquire a language based on several social-cognitive skills that they learn to use and to understand. In their interaction with other people, they understand that others also use these instruments in order to achieve a goal. Thus, nearly everything individuals in a communicative situation do is intentional. And, as mentioned above, a speaker uses a certain intonational pattern in order to (intentionally) achieve a goal, i.e. convey information in the most effective way.

To summarize, a speaker has the possibility to accent certain words or parts of an utterance in order to indicate those parts that are especially important to him. The syntactic structure of a sentence is more or less independent of the

intonational realization and gives no information about the intention a speaker has in mind when uttering a sentence¹⁷. Thus, prosody cannot be a part of any innate syntactic rules as supposed by the Nativist-Generative account. Although the Usage-Based approach does not make any specific assumptions about the role of any prosodic cues in order to achieve language, intonation seems to fit into this approach very well. First, prosody has a function that derives from its form, as proposed by the Usage-Based approach. Second, this approach assumes that children acquire a language based on several social-cognitive skills that they learn to use and to understand. As we have seen, Intonation requires these skills.

However, the question remains as to how children come to learn about the intonational conventions. To answer this question, the next chapter will give an overview of the relevant literature examining how young children get access to the (communicative) intentions of other people, followed by a brief overview of children's use of intonation when marking the informational status of referents.

4. Intonation in language acquisition

4.1. Prerequisite

As we have seen in the previous chapters, intonation is an important instrument in order to mark the cognitive status of target referents. To this end, a speaker takes into account what he assumes is part of the listener's knowledge and marks his utterance in an appropriate way with a particular intonational pattern. Additionally, the hearer needs to have the ability to understand this marking. Thus, in order to express and to understand the communicative intentions within a situation, it is essential to know how to realize and how to interpret intonation. This means that both the hearer and the speaker have to be aware of the corresponding linguistic conventions. However, this is actually the second step. In order to understand the communicative intentions of a speaker and the way this is expressed in a particular language, one has first to understand what knowledge is shared between the participants of a conversation – exactly what is the basis of their common ground.

Within a discourse, participants are developing shared common ground all the time. New entities are also constantly being introduced. In order to mark a referent as new in the discourse (because it is introduced for the first time) the

¹⁷ There are of course exceptions in which the syntactic form is an indicator of the intentional meaning. One of these exceptions is for example a cleft-sentence e.g., "It was the dog that ate the apple". However, these constructions are assumed to have a special function in the discourse, requiring a separation between logical presuppositions on the one hand and shared knowledge (as signaled by prosody) on the other. See Delin (1995) for an overview

speaker has to 'know' that this referent is in fact new within the discourse. If the speaker refers to the matter again in one of his subsequent utterances, the information has to be considered as given from the preceding discourse. Again, the speaker has to know about this givenness both within the discourse and in the mind of the listener in order to mark the referent in an appropriate way. Thus, in order to use intonation appropriately, an understanding is necessary about what other people in that communicative situation know. In particular, speakers and hearers need to know that others may have a different view of the world around them – and they need to be able to take another's perspective. In this Chapter, I will give a brief overview of the research that has addressed these two basic abilities, namely perspective taking and understanding other person's communicative intentions, which are needed in order to learn language, as described by the Usage-Based approach.

Perspective taking in infancy

Recent research provides evidence for the fact that infants of 14 to 18 months of age already understand what another person does and does not know. And, they understand that another's knowledge may be different from their own knowledge, based on previous experience.

O'Neill (1996) addressed the question of whether children understand what others know, even if that knowledge is different from their own point of view. She found that children around their second birthday not only know this, they also communicate differently depending on the parent's knowledge state. In her study, a desired object was hidden in one of two opaque containers that were out of the child's reach. To obtain this object, the child had to request help from her parent. In one condition, the parent witnessed the hiding and thus knew about the location of the hidden object. In another condition, the parent didn't know about the hiding because she had either left the room or closed his or her eyes before the hiding. Thus, the parent was ignorant of the object's location. Results suggest that children of 2–2;6 years gestured more to their parent in general and more specifically to the location of the object when the parent was ignorant of the object's location than when he or she was knowledgeable. This study shows that children know what others know because they have seen the other person witnessing an event.

In order to investigate whether 15 month old infants also have the ability to understand the underlying mental state of another's behaviour, including their expectations about the world, Onishi and Baillargeon (2005) designed a habituation study. In their study, an adult had seen an object in a certain location. However, the adult did not witness the unexpected transferal of the desired object to a new location. In the test situation, the infants had to predict that the adult would look for her desired object in its previous location. Thus, in this study infants of 15 months expected an adult to search for an object where she last saw it. In contrast, their looking-times increased when this expectation was violated; that is when they saw an actor reach for an object at its true location, which should have been unknown to the adult given that the transferal of the

object was not witnessed. Irrespective of whether this looking time study demonstrates an understanding of false belief, as the authors claim, it clearly shows that infants can keep track of what others know in the sense of what they have and have not experienced previously (see Perner & Ruffman, 2005, for an alternative explanation). In terms of language acquisition, Akhtar, Carpenter, and Tomasello (1996) addressed the question of whether young children can use the ability to take another's perspective in order to learn words. They had two year old children play with three toys successively with an experimenter and a parent. The parent then left the room and a fourth object was brought out, and the experimenter and the child played with it for the same duration as the first three. Then the parent returned and looked at all four objects, arranged in a row on a shelf, and exclaimed: "Oh, a gazzer! Wow, a gazzer! Look at the gazzer!" Children inferred that the parent wanted the object that he or she was now seeing for the first time, even though the children themselves had the same amount of experience with all four objects. Furthermore, 14 month old infants interpreted an excited reaction toward an object as meaning that it was new for the adult. However, they looked around the room for another possible referent when the intended object was not new, but was familiar to the adult (see Moll, Koring, Carpenter and Tomasello 2006). What these studies show is that young infants already have an understanding of what information another person needs in order to fulfil a certain (communicative) goal.

Understanding communicative intentions

Findings in the field of gestural communication (e.g. pointing) suggest that twelve month old infants already use pointing behaviour to communicate in an appropriate way. For example, Liszkowski and colleagues (2004, 2007a) showed that infants persisted in their communicative goal and expanded their pointing behaviour by repeated pointing and increased vocalizations when a recipient did not react to their pointing. The infants were dissatisfied when the adult's comment about a referent was unenthusiastic and therefore did not match the infant's interest compared to a situation in which the adult reacted as expected (e.g. by sharing attention and interest). And, an infant pointed more often to an interesting event when the adult had not yet seen it (Liszkowski et al., 2007b), as reflected in their differential pattern of pointing. These experimental results establish that 12-month-olds point with communicative intent. They want to refer others to specific, and sometimes even absent referents. Further research done in this field shows that infants do not only want to inform others, they even adjust their gestural pointing behaviour to the needs of a requesting adult. For example, in cases in which an adult was looking for an object, infants pointed to the location of that object in order to inform the adult about it (Liszkowski, Carpenter, Striano, & Tomasello, 2006). This happened more often when the adult was ignorant than knowledgeable of the objects' locations (Liszkowski, Carpenter, & Tomasello, 2008). What these results show is that infants know what others know—at least

in the sense that they know what objects or events others have experienced a few minutes previously.

Additionally, there is evidence for the fact that young children do not only understand others communicative intentions, they also want to be understood when communicating. For example, Shwe & Markman (1997) found that 2;0 – 2;6 old children take into account the mental component of their communicative signals. Children in this study were presented with situations in which they either did or did not get what they wanted after a request. Crucially, the experimenter either understood or misunderstood their request. The results show that children clarified their signal more often when the experimenter misunderstood their request (even when they got the toy they wanted) compared to when the experimenter understood. Regardless of whether young children achieved their goal, they tried to clarify their request to ensure their communicative act had been understood (but see Grosse, Behne, Carpenter & Tomasello, in press for an alternative explanation).

Do all of these studies, therefore show that young infants already have an understanding of the information that another person needs in order to fulfil a certain (communicative) goal? Infants, even before they have acquired language, want to convey information to another person. For example, they understand based on their own and others' experiences of an entity whether that person has seen an object before or not. And, even the youngest infants can adjust their (preverbal) communicative behaviour according to that knowledge. It seems that children fulfil the requirements that are needed in order to use intonation appropriately because perspective-taking and understanding another's communicative intentions is an essential ability that is needed in order to use intonation. The appropriate use of intonation only works when the speaker knows what the hearer knows and vice versa. Unfortunately, only a few studies take into account young children's understanding of the intention conveyed by intonation. Instead, research into children's use of intonation in recent years has mainly concentrated on how children use intonation in a linguistic sense. In the next chapter, I will give an overview of the research done in the field of children's intonational development.

4.2. Intonation in Information Marking

As we have seen in the previous chapters, the appropriate intonational realization of utterances requires strong knowledge about the cognitive status of target referents within the mind of the listener. Thus, it is of particular importance to know what the other persons within a discourse know or do not know.

One of the first to examine how young children treat elements within an intonational unit that either have or have not been previously mentioned was

Wieman (1976). She presents one of the first systematic investigations into young children's production of accentuation¹⁸ at the utterance level. Her work was inspired by anecdotal evidence in the literature, suggesting that children's stress patterns are not random, but rather are a manifestation of syntactic and/or semantic structures, as suggested by Generativists. For example, Miller & Ervin (1964) noted that one of their children (Christy) said "*CHRISTY room*" for the possessive meaning 'Christy's room', but "*Christy ROOM*" for the locative phrase 'Christy in the room'. Similarly, Bowerman (1973) reported the accent patterns of "Kendall", who in 14 out of 17 cases accented the object more heavily than the subject in subject-object phrases, and 10 out of 12 times accented the possessor in possessive phrases. However, these were only anecdotal notes and Wiemann (1974) was concerned with two questions: (1) do children in the early periods of language development use accent with any regular patterns, and (2) what are these patterns based on? She investigated five children between the ages of 1;9 and 2;5, using tape recordings of play sessions with each child. She found that the children accented the noun in adjective + noun combinations like "Blue Man", but only when it was mentioned for the first time. When it was already active and given as in "Man. Blue man", the noun was deaccented. Similar findings were also found for noun + locative combinations. Although only seven examples of this kind were found in the entire study, Wiemann suggested that children understand something about the relationships of discourse entities and operate with an appreciation of what is new in their utterances and apply stress accordingly (see also Chafe 1970). Thus, the location of the accent was not random, but was influenced by the information structure of the utterances.

In terms of the marking of the informational structure in children's language, MacWhinney & Bates (1978) were interested in cross-linguistic differences and examined how children, acquiring one of three languages (English, Hungarian & Italian), mark elements that vary along the pragmatic dimension of givenness vs. newness. They asked 3, 4 and 5 year old children to describe triplets of pictures, in which certain referents increased in givenness. For example, one series of three pictures showed a boy doing three different actions e.g., "A boy is running / skiing / swimming". In this example, the pragmatic status of the subject increases in Givenness whereas the status of the verb increases in Newness. The authors analyzed accentuation, amongst other linguistic properties like ellipsis, pronominalization and (in)definite article. They found a main effect for accentuation on that element that increased with newness, especially from the English learning children. Additionally, older children used more accentuation as opposed to the younger age-group. However, this was not statistically significant

¹⁸ Several authors presented in this Chapter used the term 'stress' within their studies to describe prominence at both the level of word and utterance. The usage of this term was commonly accepted for all kinds of prominence. However, for the sake of consistency, I will continue with the distinction between stress (for prominence at word level) and accent (for prominence at utterance level), as described in Chapter 2.2., unless otherwise indicated.

and the authors concluded that the use of accentuation has already been acquired by the age of three.

The aforementioned studies present research conducted in order to answer the question of how different referents are realized in young children's speech with respect to their status in the interlocutor's mind. In addition to this, several researchers investigated how children realize the pragmatic functions of intonation within an utterance, e.g. accenting certain words or phrases within an intonation unit in order to mark it as the most informative part. For example, Hornby & Haas (1970) investigated the use of 'emphatic stress' in a situation in which there was a referential contrast between different actors or events. English-learning preschool children at the age of 4 years were asked to describe pairs of pictures in which either the actor or the action changed (e.g., 'a boy riding a bicycle' vs. 'a girl riding a bicycle'; and 'a man washing a car' vs. 'a man driving a car'). The results of this study clearly suggest that children at the age of 4 years accented the newly introduced referent in the second picture. (see Baltaxe (1984) for similar results with 3-4 year old children in comparison with autistic and aphasic children).

Overall, it is unclear how "emphatic stress" was defined in this study. Accordingly elements "were scored for contrastive stress" (1970:397). Additionally, in the MacWhinney & Bates (1987) study, the coding procedure was described as follows: "elements were judged to be emphatically stressed if they [...] received more intonational stress than any other item in the response" and if "the amount of stress was decidedly more than would be given in a neutral [...] rendition of the utterance" (1978:548). However, the exact form of the accent or its prosodic features is not clearly defined. This is consistent with developmental studies of the time, that often conflate "stress" and "accent", as already described in Chapter 2.2. Additionally, accentuation was mainly measured on the basis of auditory impressions, which is in itself not problematic as it reflects the common practice within the tradition of the time. But, over the past ten years or so, more advanced technology has been developed which allows more detailed and systematic acoustic measurements, including e.g. duration, pitch range and amplitude.

A recent study that investigated how German-learning 4 year old children and adults realize the intonational marking of referents in new (and thus focused) position was done by Müller, Höhle, Schmitz & Weissenborn (2006). They used an imitation task, in which short comic strips consisting of three pictures were presented. The relevant contextual information was presented orally by an experimenter, followed by a question-answer pair related to the last picture of the sequence. This last sequence was the target element and should be imitated. Interestingly, the auditory material of the presentation of this question-answer pair was systematically manipulated such that no information about any focus-related prosodic information in the target sentences was provided (the F₀-value for each word of the sentence was set to 150 Hz). However, this target sequence was to be repeated by the participants. All sentences consisted of a subject, a direct object and a verb. The sentences differed with respect to their constituent order and with respect to the focused constituent. Half of the sentences were

syntactically canonical (subject-verb-object) and the other half was syntactically non-canonical (object-verb-subject). The subject was the focused constituent in half of the sentences while it was the object in the other half. However, it was assumed that their realization would carry a natural prosodic realization. An example display of the target sentences is presented in Figure 4:

	Syntactically canonical	Syntactically non-canonical
Subject-Focus	<p>EVA (subject) kauft die Gurke (object).</p> <p>EVA (subject) buys the cucumber (object).</p>	<p>Den Opa (object) grüsst PETER (subject).</p> <p>The grandfather (object) greets PETER (subject).</p>
Object-Focus	<p>EVA (subject) putzt die FENSTER (object).</p> <p>Eva (subject) cleans the WINDOWS (object).</p>	<p>Die TREPPE (object) fegt Peter (subject).</p> <p>The STAIRCASE (object) sweeps Peter (subject).</p>

Figure 4: Experimental conditions in Müller, Höhle, Schmitz & Weissenborn (2006). The target referents in focus position are printed in bold.

The authors found that in the utterances of German 4-year-olds a focused element carries a higher pitch than an unfocused element with the same syntactic function and the same position within an utterance. This was similar to the results of an adult control group. In addition, both groups realized the initial constituent of the utterances with a higher pitch than the final one, irrespective of being focused or not. A second main finding was the strong tendency for the production of sentences with canonical word order: the children as well as the adults show a tendency to produce canonical word order (SVO) irrespective of whether the subject or the object is being focused. The authors concluded that the mastery of the prosodic devices of focus-marking is acquired early in life (as already suggested for English by MacWhinney & Bates, 1978 and Hornby & Hass, 1979). Additionally, and more important for this thesis, children had an understanding of which constituent was in focus and thus, which element in the utterance would be more appropriate to receive an accent. What this means is that they use linguistic means to express the relevant aspects of information structure.

Chen (2007) conducted a similar study, but she was more concerned with the question of what kind of pitch accents children used. She employed more sensitive acoustic measurements in order to analyze the intonational realizations according to the Dutch TObl system. In an imitation study, she examined how Dutch-acquiring preschool children at the age of 4-5 years use different pitch accent types and deaccentuation to mark the pragmatic function (topic & focus) of target referents and how this realization differs from adult's intonational behaviour. Additionally, topic and focus were counterbalanced with respect to

their position within the sentence (initial and final). Chen presented thirty-six question-answer pairs as the experimental stimuli. All answers were SVO sentences in which subjects and objects were realized as full Nominal Phrases (NP's). Within this method, the two variables were controlled for in the answer sentences: the pragmatic condition (referents were either topic or focus) and the sentence position condition (either initial or final). Half of the question-answer pairs represented the initial focus-final topic condition and the other half represented the final focus-initial topic condition. Each sentence-initial NP and each sentence-final NP occurred in both groups of answer sentences but in different combinations so that each answer sentence was heard only once. An example of the conditions in Chen (2007) is given in Figure 5.

	sentence initial	sentence final
FOCUS	Q: (Who protects the forest?) A: The FOX protects <u>the forest</u> !	Q: (What does the toad eat?) A: <u>The toad</u> eats a berry !
<u>Topic</u>	Q: (What does the toad eat?) A: <u>The toad</u> eats a berry !	Q: (Who protects the forest?) A: The FOX protects <u>the forest</u> !

Figure 5: Experimental conditions in Chen (2007). The target referent in focus position is printed in bold, the target referent in topic position is underlined.

The results clearly show that children realize both referents that stand in topic and focus position with a similar level of frequency with the H*L pitch accent. This is somewhat different to the results of an adult control group which shows that adults on the one hand realize referents in focus-position typically with the H*L pitch accent, independent of sentence position. On the other hand, the intonational realization of referents in topic-position differs regarding their sentence position. Whereas topic in sentence-final position was typically deaccented, sentence-initial topic was, like focus, mostly realized with the H*L pitch accent. And, as opposed to adults, children frequently realize the topic with an accent. What this shows is that Dutch-acquiring 4- to 5-year olds, as adults, use intonation to realize full NP topic and focus. To do so, both adults and children use the same types of pitch accents to mark the topic-focus distinction, though children's repertoire of accent types is different to those of the adults. Like adults, children deaccented the topic more frequently than the focus independent of sentence position. And, children accent the focus more frequently than the topic. This is important because it shows that children are sensitive to the use of intonation in order to realize different parts of a sentence and to distinguish between their different informative roles. The fact that children do not distinguish between sentence positions shows that they do not consider a special sentence position to have a special pragmatic role, which stands in sharp contrast to earlier studies (e.g. Hornby & Haas, 1970; MacWhinney & Bates 1978).

Arnold (2008) links the question of the pragmatic function of intonation to the mental representation of discourse referents in the mind of the listener. In her comprehension study, Arnold wanted to find out whether preschoolers use the preceding discourse context to guide their initial interpretation of referring expressions. In order to understand children's pragmatic abilities to understand the status of discourse entities based on their intonational realization, two research questions were combined: (1) do children understand different degrees of accessibility between two critical objects (e.g., bacon/bagel), when only one of them was mentioned before and the other was completely new in the discourse, and (2) how does children's use of accentuation during on-line reference comprehension work, tested by measuring eye-movements. Using an object-moving task, different pictures were presented. The objects on these pictures represented cohort competitors, meaning that the initial segment of both referents were similar, as in 'bagel' / 'bacon' or 'candle' / 'candy'. The participants received instructions for each visual stimulus, e.g. "Put the bacon on the star. Now put the bacon (alternatively: bagel) on the square". The object in the second instruction was the referring expression of interest, e.g. "bacon" in this example. The other object with an overlapping name (e.g., the bagel) was the competitor. The first instruction mentioned either the target (the anaphoric condition) or the competitor (the nonanaphoric condition). The target referring expression was either accented or unaccented. The auditory instructions were pre-recorded and manipulated so that, in the accented condition, the target word carried a pitch accent which was acoustically prominent and relatively long. In all accented conditions, the target word was realized with a L+H* pitch accent, followed by an L-H% boundary tone, resulting in a prominent sounding accent. In the unaccented condition, the target word carried no pitch accent, and was acoustically attenuated, with a shorter duration, and no boundary tone. Thus, the focus of this study lies on the different acoustic properties of the target word. Results from this study suggested that 4 and 5 year old children respond differently to accented and unaccented tokens during spoken reference comprehension. Similar to adults, unaccented words led children to initially look at the previously-mentioned object. When an unaccented word referred to the unmentioned object, children erroneously treated the word as if it were anaphoric. By contrast, in the accented condition they showed no early preference for either previously-mentioned or new referents. The contrast between accented and unaccented expressions emerged on the children's first look after hearing the beginning of the target word. This suggests that accenting – or the lack of it -- does guide children's initial hypotheses about what a word refers to. At the same time, Arnold found that children are not fully adult-like in their use of accenting. First, eye movements in response to the target word occurred later in time for children than adults. And, adults differentiated more robustly between accented and unaccented expressions. However, the overall picture suggests that 4 and 5 year old children are able to use accentuation during their on-line interpretation of referential expressions, even if they are not yet fully adult-like.

Because Arnolds study concentrates on children's comprehension of intonation, it leaves open the question of how children realize the informational

status of referents within a more complex discourse situation and how this is done when accessible referents are not visually accessible for the speaker and the hearer. Additionally, information about the type of pitch accents with which the children realize the informational status of target referents is missing.

DeRuiter (2010) tries to fill this gap by asking the question of whether children use the same pitch accents as adults and whether their use of different pitch accents changes with age. In her study, deRuiter used a picture story-telling task in order to elicit natural data. She asked children at the age of 5 and 7 years to describe picture books, in which one of four target referents varied in their informational status over the discourse of that picture book. The status was either new (the target referent occurred for the first time), given (the target referent occurred immediately after the “new - condition” picture) or accessible (the target referent re-occurred within a certain distance of the “new-condition” picture). She found that both age groups have in fact learned to mark information status by intonation. And, they do this in an adult-like way because newness was realized with an accent and givenness with lack of accent (this is in line with current literature, e.g. Baumann, 2006). Interestingly, the children do not treat every referent that has already been mentioned as given. Instead, accessible referents were realized in a way that was similar to new ones, resembling an adult’s behaviour. What this shows is that children of this age are in fact sensitive to the status of target referents within a discourse – and they use intonation to mark this. The only difference from adults was the type of pitch accents that was used in order to realize accessible referents. Whereas the children used the L+H* pitch accent more often for new referents, adults marked the accessible referents with this pitch accent. However, although children’s use of pitch accent type seems to be similar to that of adults, children appear to differ from the adults in the use of other pragmatic and para-linguistic features of intonation. For example, the children did not use any continuation intonation. They did not use the typical phrase-final rising intonation in order to indicate that the speaker is about to say more (also known as the ‘turn-taking’ device e.g., Couper-Kuhlen & Selting, 1996). But, deRuiter found a significant age difference in the functional approach to intonational realization: the older age group used it to some degree and therefore different properties of intonation seem to develop over time. Additionally, children’s use of the same pitch accents as adults does not mean that the children do not have to learn more about the phonetic realization of the different intonational contours. For example, children in this study produced accents with smaller excursion and flatter slopes than adults. And, adults realized the pitch minimum earlier and the pitch maximum later within the words than children.

On the whole, the studies reviewed in this section show to some extent that young children do understand that different cognitive states of referents within a discourse are marked in different ways, depending on the context and the degree of givenness of the target referents. However, as already mentioned, studies from the 70`s and early 80`s are difficult to interpret. In these studies, it is not really clear what was measured and how. For example, in Wieman’s (1974) study, the relative prominence of words was mainly investigated within one

intonational unit. But, in order to investigate the cognitive status of target referents and its relation to the overall cognitive abilities, the intonational realization of words and/or phrases can only be interpreted as related to the overall discourse. This means that in order to understand anything about young children's intonational behaviour, it is not just the individual realization of a word that needs to be taken into account, but rather the overall intonational behaviour within a situation or a linguistic discourse. Thus, anecdotal evidence in which an infant uses one of several possible intonational contours in any situation seems to be an inappropriate measure for infants and young children's intonational development. Additionally, the general cognitive development has to be factored in, as intonation is part of the overall discourse situation. Thus, the prosodic cues that mark the relative importance of words can only be interpreted meaningfully when the discourse context in which they are embedded is considered.

More recent studies take into account children's phonetic realizations, measured with more sophisticated methods. But, these studies have mainly concentrated on the linguistic part of intonation and its role within an utterance. Thus, examination of children's intonational marking of the focus (what is important) and the background (what is less important) are methodologically well defined. But they do not answer the question of whether children really understand another's cognitive status of target referents within a discourse based on the intonational realization; in other words, whether they understand what another person is referring to. Furthermore, most of the studies presented in this section test how children realize an utterance in cases in which something is new or given for both the speaker and the hearer. In natural conversation, this does not work like this because the speaker knows things the hearer does not know. The speaker has to take this into account and adjust his intonational behaviour with respect to this. Studies testing this aspect (see e.g. deRuiter, 2010) concentrate mainly on older children that are already exposed to language.

To summarize, it is unclear how intonation affects young children's ability to understand what another person is referring to. And, it is unclear whether young children can use intonation to understand intentions and thus, to acquire language. But, this is an important element for understanding young children's cognitive development. As we have seen in Chapter 3.2., the understanding of intentions is essential for acquiring the social-pragmatic and cognitive skills that are needed to learn language. And, intonation does convey information about the informational status of elements within an utterance. Thus, in order to understand the language acquisition process, young children's competence in the area of intonation, both in production and comprehension, has to be taken into account.

5. Research questions

This short literature review shows that the intonational marking of information has attracted a great deal of attention. However, investigating children's pragmatic use of intonation is a challenging task because it is strongly related to their overall pragmatic and social-cognitive abilities. In order to comprehend and to realize an utterance correctly, both the speaker and the hearer have to be aware of what information is and is not important. Prosodic cues allow the listener to interpret the relative importance of each word or part of the utterance and to represent the informational status of discourse entities accordingly. As we have seen, it is of particular importance to understand another's perspective when acquiring a language as well as when using a particular intonational pattern. What this means is that the acquisition of language requires a certain mind-reading ability and intonation deals exactly with this point. Our understanding of intonation potentially plays a crucial part in the acquisition of our broad social-cognitive abilities, which are influenced and extended by it. However, with respect to (first) language acquisition, studies examining whether young children understand the intention conveyed by intonation are rare. Reviewing evidence for young children's overall pragmatic and social-cognitive abilities, however, it seems plausible that they have sophisticated abilities they could use to understand the intentions of others, taking intonation into account. But, to my knowledge, this has never been directly tested. Instead, recent studies have mainly investigated the role of intonation in children's interpretation of the information structure of sentences; that is "what is the sentence about" or "what can the sentence be contrasted with from a logical perspective?" In order to fill this gap and to investigate the question of whether children acquiring a language can use intonation in order to understand another's intention, the following questions will be addressed in this thesis:

- When and how do children develop an understanding of the possibility of realizing intentions by intonation?
- Can children use the intonational cue in order to find out what another person is referring to and, related to this, can they use this knowledge to learn language?
- Can this knowledge pave the way for the acquisition of more complex, syntactic constructions?

These questions will be addressed in Part II as they deal with children's ability to comprehend another's intention by way of intonational realization.

Additionally, Part III will present empirical evidence about young children's productive behaviour when realizing the informational status of target referents within a discourse. In particular, the questions addressed in this section are:

- Do young children use intonation to realize the cognitive status of target referents within a discourse?
- What role does the input play in the acquisition of intonation?

Part II: Empirical Studies - Comprehension

6. Referential function of intonation

6.1. Understanding intentions by intonation

6.1.1. Introduction

There are two basic ways in which adults draw young children's attention to particular objects in the environment: by pointing (and other deictic gestures) and through using words (and other linguistic conventions). Comprehension of pointing gestures seems more instinctive because it is based on infants' (and other primates') natural tendency to follow another's gaze direction to external targets; an ability that is mastered from the age of six months (Moore & D'Entremont, 2001, cf. Chapter 4.1.). Typically, infants will begin to point before they use language (Carpenter, Nagel & Tomasello, 1998). What makes pointing, and other deictic gestures, so natural and pragmatic is the fact that they direct another's visual attention to an object or event in the here and now. Words and other linguistic expressions, on the other hand, are more conventionalized and become effective only through the social learning of a convention. For example, all users of a communicative system have to use the same 'arbitrary' sound for the same referent in the same way to direct attention, typically, to a particular kind of referent. Common nouns and most verbs within this communicative system are not used to refer to particular objects or events, as is the case with pointing; that is, not without some kind of grounding device, such as determiners or tense markers. Instead they refer to classes of particular kinds of objects or events. This is important for the language learning process. In order to learn a new word, children need some kind of independent social-pragmatic information about what the adult is referring to when using a new word – and pointing is a particularly effective source of such independent information (e.g. Tomasello, 2001, 2008). In general, a growing body of research suggests that children's word learning rests fundamentally on their social-pragmatic skills, within which an understanding of the pointing gesture plays an important role (e.g. Baldwin & Moses, 2001; Saylor, Sabbagh & Baldwin, 2002; Saylor, Baldwin & Sabbagh, 2004; Tomasello, 1992, 2003).

Another, indirect cue that children can use in order to find out what another person is referring to is the knowledge of what another individual regards as given and new. As we have seen in Chapter 4.1., recent research has found that even the youngest infants already have this ability. Additionally, several studies have demonstrated that children are aware that an adult's focus of attention may be different from their own, and this is supported by studies showing that children are able to use a variety of cues to determine an adult's focus (Akhtar & Tomasello, 1996, Tomasello & Barton, 1994), especially during joint engagement (Moll & Tomasello, 2007). However, most of these studies do not attempt to control for intonational patterns. Instead, these studies investigated the psychological perspective and identify a variety of mechanisms that children rely on when inferring the meaning of words. To do so, children were exposed to

a situation in which an adult either did or did not witness a particular event (e.g. hiding an object or playing with an object) and they concentrate on whether children understand what the other person does or does not know when requesting that object. For example, Tomasello and Haberl (2003) had infants of 12 and 18 months of age play with an adult and two novel toys successively. The adult left the room before a third toy was brought out by an assistant. During the adult's absence, the infant and assistant played with the third toy. Finally, all three toys were held in front of the infant, at which point the adult returned to the room, exclaimed excitedly, then produced an unspecified request for the infant to give her a toy (without indicating by gazing or pointing which specific toy she was attending to). Surprisingly, infants of both ages selected the intended object because it was new for her.

In order to solve this task, infants had to understand what the adult knew and did not know in the specific sense of what she had and had not experienced previously. This is a remarkable skill given that an understanding of the knowledge-ignorance distinction had previously only been shown for toddlers over 2 years of age (see e.g. O'Neill, 1996). As shown in the previous chapter when looking at the research conducted by Akhtar, Carpenter, and Tomasello (1996), the acquisition of language is related to the question of whether children use knowledge about givenness and newness of objects in order to learn new words. These authors showed that children know what objects or events others have previously experienced. And, children can use this knowledge to learn the word for a particular object. When the requesting adult gave the particular object a name, children of 24 months learned the name for this object. What this shows is that children can use novelty from the discourse context in order to learn new words (but see Samuelson & Smith, 1998; and Diesendruck, Markson, Akhtar & Reudor, 2004 for an alternative interpretation).

However, in the test situation, the request was not controlled with regard to its intonational realization. Instead, as Moll and Tomasello (2007) report, they "exclaimed [the object] in a tone of excitement" (Moll & Tomasello, 2007:312). This is a very natural manifestation as it is well known that mothers, when talking to their children use intonation to highlight important linguistic information such as labels for unfamiliar objects both in pitch and duration (e.g. Saffran et al. 1996). New words tend to appear at points of perceptual prominence both in place and frequency, even at the expense of grammar violations (Fernald & Mazzie 1991, Aslin 1993). When adults speak to children they use higher fundamental frequency, wider F0-excursions, shorter utterances, longer pauses, slower articulation and more prosodic repetition in speech that is directed to children, as opposed to adult-directed speech (e.g. Fernald & Simon 1984, Papousek et al. 1987). And even vowel lengthening is more exaggerated to mark both phrase and clause boundaries (e.g. Morgan 1986). Moreover, infants prefer listening to this speech style, even when spoken by strangers (Fernald 1985) whereby F0 is the primary acoustic determinant (Fernald & Kuhl 1987).¹⁹

¹⁹ We will come back to the characteristics of CDS in Chapter 9

However, as part of this excitement, several studies also report about lexical items that were used in order to mark their request. For example, Moll, Carpenter & Tomasello (2007) reported that they “excitedly exclaimed ‘Oh, look! Look there! Look at that one there!’, which the experimenter followed immediately with the request ‘Give it to me, please!’ (2007:4)”. Similarly, Tomasello & Haberl (2003) reported that they used lexical items like “Oh, wow! That’s so cool! Can you give it to me?” while the experimenter was gesturing ambiguously in the direction of the objects. What this means is that in all these studies, many different cues (e.g. lexical items, hand gestures, facial expressions) were used in order to make their request clear and children could use all of these cues, i.e. the whole “package”, in order to find out what that person is referring to.

The question then arises, what role intonation does play in this package of communicating surprise (about a new and unexpected entity)? Surprise is biologically combined with a certain bodily expression (e.g. Ekman 1984, 1999) and “all emotions are expressed through both physiological changes and stereotyped motor responses” (Plutchik 2001:344). Related to this, Gussenhoven’s ‘Effort Code’ (cf. Chapter 2.3.2.) explains that increases in the effort of a certain communicative act results in greater articulatory precision and in wider excursion of pitch movements. Pragmatically, speakers exploit this fact to convey a certain meaning. This meaning can be derived from the effect of the expenditure of effort, i.e. the speaker is being forceful because he thinks that the information, conveyed by his message, is important.

Thus, the question is whether young word-learning children can use the intention another person, conveyed by intonation in order to find out what another person is referring to and whether they can use this in order to learn new words. Grassmann & Tomasello (2007) tested whether the prosodic characteristic of child-directed speech facilitates children’s word learning and whether children learn a novel word for novel referents or if prosodic highlighting of novel word plays a role. In this study, the authors demonstrated that 2-year-olds only relied on discourse newness in their interpretation of a novel word, when the novel word was accented. In one of their conditions, a nameless novel object was new to the situation and in another condition a nameless action was new to the naming situation. Children heard the experimenter say two novel words in an intransitive sentence, a novel verb (‘miekt’) as well as a novel noun (‘feks’): ‘Der Feks miekt’. As a second factor, sentence accent was varied: either the novel noun or the novel verb was accented. The results revealed that children learned the novel noun (Feks) for the novel object only when the noun was accented and the novel object was new in the situation but not when the noun was accented and the novel object was given. Grassmann and Tomasello (2007) suggested that this indicates that children interpret sentence accent in language as being iconic of the speaker’s intention to refer to a salient aspect of the situation.

In a related study, Grassmann & Tomasello (2010) investigated 24-month-olds’ comprehension of prosodic stress using a looking-time measurement. In particular, they wanted to know whether children focus their visual attention on new referents when the corresponding word is stressed in an utterance. To do so, the children saw pictures of highly familiar objects (e.g., a ball). In a second

picture, containing two highly familiar objects, one of these objects was the same as in the first picture (e.g., the ball), and thus was an established referent ('given' information), while the other object was new (e.g., a dog). However, before the second picture was revealed, the children heard a sentence such as "The dog has a ball" – where the stress fell either on 'dog' or on 'ball'. The results indicate that children did focus their visual attention on the referent of a familiar word when the word was accented and the referent was new to the situation. Importantly, neither accentuation on a word nor newness of a referent alone led the children to visually focus on the corresponding element (i.e., the referent of the acoustically salient word or the new element in the situation). What this shows is that children assume that the acoustic salience of words is related to the contextual salience of the referents. This supports the assumption that children understand that the prosodic salience of a word has something to do with the intention of a speaker, namely to direct attention to something that is new in a situation.

Although deRuiter (2010) (cf. Chapter 4.2.) found that children at the age of 5 begin to use intonation to signal the informational status of discourse referents, there is to my knowledge no study about the comprehension of the intonational realisation of given information in contrast to new information in a discourse context. However, it remains unclear whether accenting – or the lack of it – guides children's initial interpretation about what a word refers to and whether young children use the connection between the knowledge of what another does and does not know and the corresponding prosodic markings of the informational status of referents to learn new words. Thus, the question is: Do children understand that a speaker has a certain intention when using a certain intonational pattern. Related to this, the question arises whether children already have knowledge of the linguistic convention concerning typical newness and givenness accents; that is, do they *understand* the intention behind the use of different intonational realizations of discourse referents to mark their state in a preceding discourse?

In the current study, therefore, I systematically manipulated the factors newness versus givenness of objects, depending on whether or not an experimenter had seen one of three objects before. I tested 20-month-old children using a method similar to Tomasello & Haberl (2003). After the children had seen an experimenter either witness an object or not, I wanted to know which object they would hand over when the experimenter ambiguously requested one of these objects. What is new in the present study is that the request for one of the objects no longer consists of a whole package of cues. Instead, the request is only marked by intonation, either with the Newness – accent H* or with the Givenness – accent L*. My prediction was that the pitch accent used in the givenness condition would lead children to choose the third object, which was new to the speaker less often than in the newness condition.

6.1.2. Data & Method

Following Tomasello & Haberl (2003), I used an object-choice task to evaluate 20-months-old German children's understanding of intonation as a cue to the intention of a speaker. Additionally, I wanted to investigate whether children at this age can distinguish between different types of accents. To do so, I presented three novel objects, two of them were witnessed by an experimenter while the third one was not. After this, the experimenter ambiguously requested one of the three objects by marking his request with either the newness – accent H* (indicating that he is surprised and is requesting the new object, which has not been seen before) or with the givenness accent L* (indicating that he is not surprised and is requesting one of the objects he has seen before). To make sure that the intonational pattern of the request was consistent throughout this study, the utterances were performed by a GToBI-trained experimenter.

Participants

The participants of this study were obtained from a database of parents from a middle-sized German city who had volunteered for studies of child development. Participants were 60 (28 females, 32 males) monolingual German 20-month-old children (mean = 20,1 month, range = 19,2 – 20,6). An additional 15 children were tested but had to be excluded from the final sample, for one of the following reasons: they failed the warm-up task (N= 7), because of experimenter error (N= 4), because of uncooperativeness (N= 3) or because of bilingualism (N=1).

Materials and design

In order to find out whether 20-month-old children understand the intention behind a certain intonational pattern based on the speakers knowledge, two experimental between-subjects condition were created. The children's task was to identify the referent of a novel target word. In order to identify the correct referent of that target word, the children had two cues: their knowledge about what the requesting person knew about the different objects and the intonational pattern of a request. The word used in both conditions was a phonotactically correct disyllabic German pseudo-word (`Flomer`) which was embedded in a typical and appropriate German request. The main difference between conditions was the kind of accent used during the request: In the newness condition I used the typical marker for contextually new referents (H*), in the givenness condition the referent of the novel object was marked with the appropriate marker for given referents (L*).

For the experimental test, three novel objects were created. These were either hand - made or hardware items that children of this age were unlikely to know (see Figure 6).



Figure 6: novel objects used in this study. (A) shows a modified bird-cage mirror, (B) a modified card-holder and (C) a modified salt-jar

Each of the novel objects was a different color and shape, but were approximately the same size. A special move was assigned to each and as a consequence they were manipulative in a particular way. The playing procedure with each toy followed a standardized script, which was identical across conditions and toys. A pre-test for children's preference ensured that all novel objects were equally interesting to children of that age. The order for the two conditions as well as the order in which the toys were presented (first, second, third) and the toys' location on the tray in the response phase (left, middle, and right) was counterbalanced. Each child was randomly assigned to one of the two conditions, yielding 30 children in each condition (mean age in each condition: newness = 624 days, givenness = 625 days).

Procedure

Participants visited a child laboratory with a parent for one session lasting approximately 15–20 min. The parent never engaged in the interaction. Prior to the study, the experimenter (E1) and an assistant (E2) played with each child in a playroom until the child was comfortable with the situation. The experiment took place in a testing room (4.30 x 4.30 m) on a square table. The child was positioned on the parent's lap and sat 90° from E2 and 180° from E1, who was seated with his back to the door.

Warm Up: A warm-up task was conducted in order to see whether the child understood the object choice test and whether the child was able to react to E1's request. The experimenter placed three familiar objects on a tray. Following this, E1 asked for each of them one by one using their names. The objects were three familiar animal-toys: a cow, a dog and a cat. To pass the warm-up task, the child had to hand over either the first or the second requested object.

Test Trial: At the beginning of the experimental test, E2 brought out the first toy, showed it to the child and E1, saying: "Look what I have!" She then demonstrated how to manipulate the object such that it would make a certain move. The child and E1 then took turns manipulating it. During this time, E1 and E2 commented on the joint action in a very general fashion, saying, "Look at what you can do

with this!” and “That’s nice!” None of the novel objects was labeled during the play but pronouns were used (the German equivalent for ‘it’, ‘this’, or ‘that’). After 40 s, E2 took the toy and placed it on a tray out of the sight of the children, saying, “I’ll put this here!” She then brought out the second toy, and exactly the same procedure was repeated for this toy.

Before the third toy came out, E1 left the room using the pretext of a telephone call. He stood up, waved to the child and to E2, saying: “Bye, Child, Bye E2”. After he was gone, E2 advised the child that E1 was out of the room and could not see or hear them but that they would play with another toy. She then brought out the third toy and repeated the same procedure as for toy 1 and 2. After they had finished playing with the third toy, E2 took the tray with the toys on it and put it on the edge of the table. She then put an additional, empty tray opposite the child. Both trays were out of the child’s reach. She began to move each of the objects from the first tray onto the empty tray saying: “I’ll put this here!” She moved the objects in a counterbalanced order and all utterances were realized with the same intonational pattern. In doing this, the child once again had the opportunity to watch all of the toys²⁰. By using neutral intonation, none of the toys received special emphasis. E1 then came back into the room and said: „Hello, I’m back“. E1 remained in front of the table at a distance of approximately 1 m. At that moment, E2 held the tray with the toys on it straight in front of the child, so that all objects were equidistant from the participant. E1 watched the toys for approximately 3 sec., then said: “Ah, Child, give me the *Flomet!*” The intonational realization of the request was dependent on the condition. He then approached the table and held out his hand to enforce his request. In order not to provide the child with any cue, he held his hand toward the middle of the tray at an appropriate distance and looked the child in the eyes. He repeated his request up to two times if necessary. Figure 7 summarizes the procedure.

²⁰ The reason for having two trays was so that all of the objects would be present for the same amount of time.

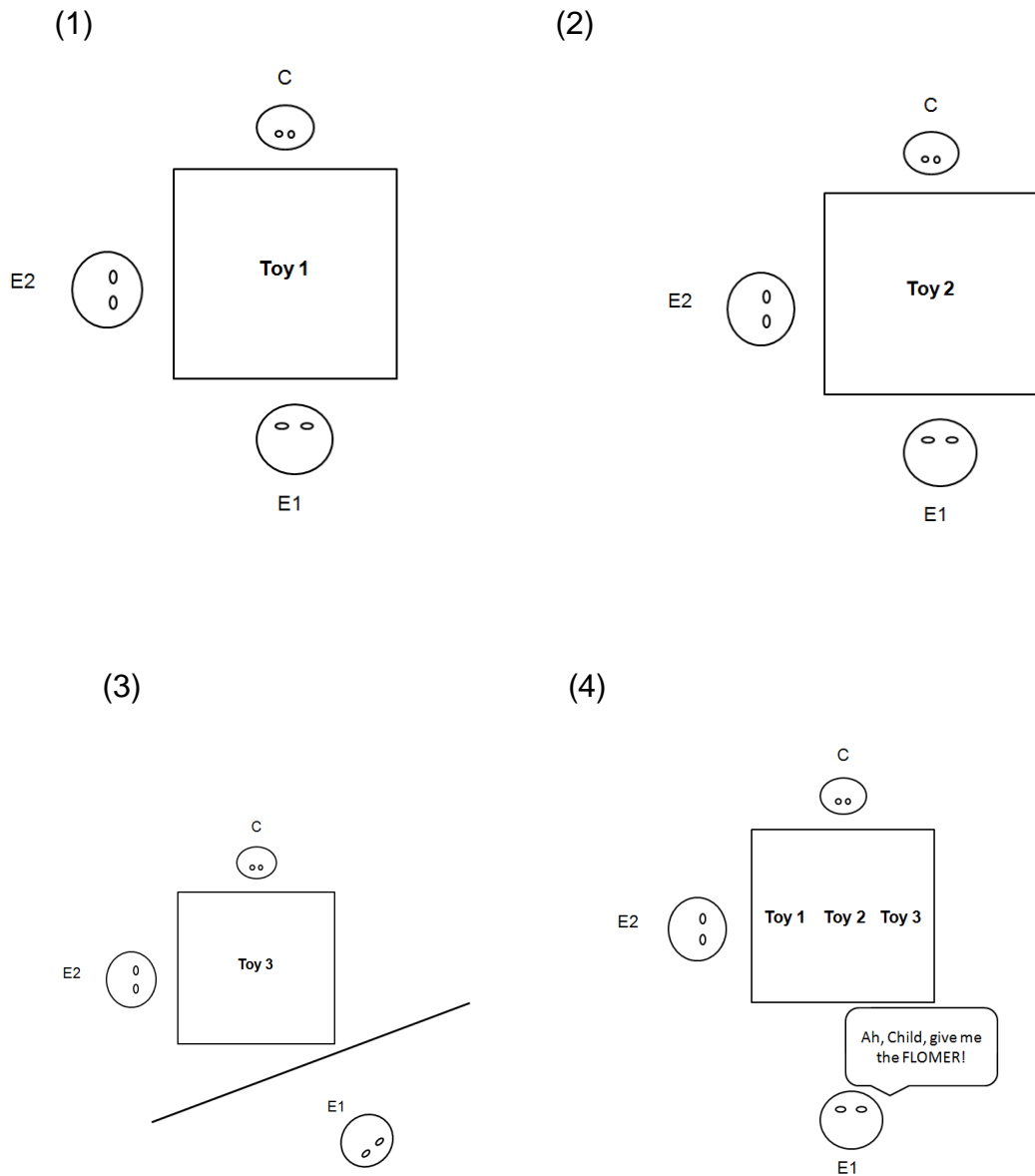


Figure 7: schematic summary of the procedure. E1, E2 and the child play with two toys consecutively for about 40 seconds (1) & (2). Subsequently, E2 and C play with a third toy while E1 is not in the room (3). After playing with all three toys, E2 puts all of them onto a tray on the table. E1 comes back into the room and requests an object, using a nonsense-verb (4).

Acoustic properties of the test material

In the newness condition, the intonational realization of the nonsense word 'Flomer' was marked with an H* with a preceding rise, high fundamental frequency, wide F0-excursions, expanded duration of utterance and pauses and a lower speed of articulation²¹ (see Figure 8).

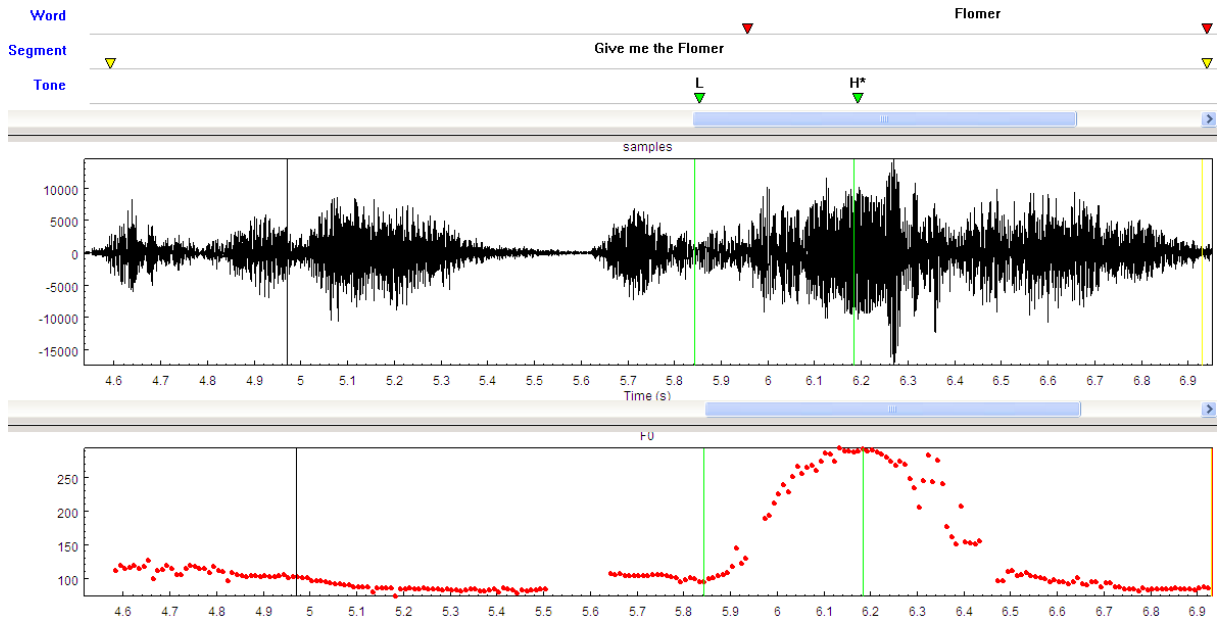


Figure 8: Intonational realization of the test-utterance in the newness Condition. The first row shows the text level, the second row shows one the oszillographic representation. The third row represents the intonational contour of the utterance given a sharp rise up to the F0 peak, indicating a leading low tone, making it a L+H* on the target word 'Flomer'

In the givenness condition the intonational realization of the nonsense word 'Flomer' was marked by a L* pitch accent, characterized by lower fundamental frequency, narrower F0-excursions, shorter duration of utterance and pauses and a higher speed of articulation (see Figure 9).

²¹ It is important to note that child-directed speech tends to be more slowly articulated than adult-directed speech (Garnica, 1977). In the word-learning process, this leads to more clearly articulated vowels so that their vowel categories overlap less in formant characteristics (Bernstein-Ratner, 1985).

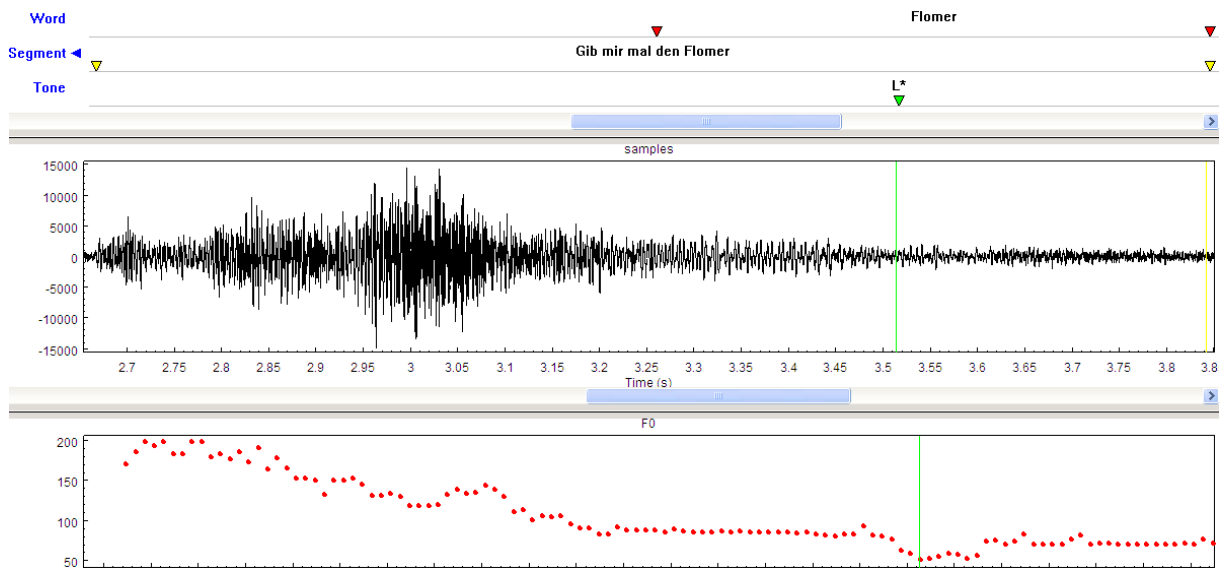


Figure 9: Intonational realization of the test-utterance in the givenness Condition. The first row shows the text level, the second the oszillographic representation. The third row represents the intonational contour of the utterance with the L* accent on the target word 'Flomer'

The acoustic speech signal of the request was analyzed for the length of the utterance and the target word as well as the mean time at which the pitch accent reached its peak within the word. Additionally, the mean frequency of the pitch accent was measured. The request in the givenness condition was marked by a flat contour with a low pitch accent at 73 Hertz, whereas the intonation contour for the referent in the newness condition was characterized by a rise from about 134 Hertz to 283 Hertz. The difference between the high target point and the preceding low beginning correspond to an average difference of 14 semitones. Furthermore, the pitch accent in the givenness Condition was realized earlier than in the newness condition. The distinction between the requests for a new respective given referent was predominantly realized by a greater F0-excursion and a different kind of F0-contour and pitch accents, but also by a different length of request. This was obtained by slower articulation in the newness-Condition, but also by longer pauses between the words, especially before the target referent 'Flomer'. The following table summarizes the acoustic properties of the target words and utterances in the two conditions.

Table 5: acoustic properties of the target utterance and the target word in both the newness and the givenness Condition

	mean length (utterance)	mean length (target word)	Mean time pitch accent (within target word)	Mean F0 (target word)	
				low	high
Newness Condition	2538,14 ms (88,82)	1236,77ms (56,95)	473,15 ms (44,17)	134 Hz (52)	283 Hz (28)
Givenness Condition	1292,47 ms (88,72)	684,88 ms (87,76)	324,29 ms (68,22)	73 Hz (9)	----

Coding and reliability

The first experimenter did a live coding and judged which of the three objects the child handed over. Additionally, the test sessions were recorded which made it possible to do a control coding immediately after the session. To assess inter-rater reliability, a research assistant, who was unaware of condition, coded 20 % (12 participants) of the final sample from the video material. Because of one disagreement between the first and the second coder, which turned out to be an inadvertent mistake, all final samples were checked once more so that the agreement between the two raters was 100%, for a Cohen's kappa of 1. In addition, 50% (15 new, 15 given requests) of the intonational realization of the request "Give me the Flomer!", was tested by a blind coder and compared to speaker's intention during the test-phase²². Agreement between the two raters concerning the intonational intention was 100%, leading to a Cohen's kappa of 1.

6.1.3. Results & Discussion

Figure 10 shows the number of children's object choices separately for the two conditions, with 'Toy 1', 'Toy 2' and 'Toy 3' referring to the temporal position of the toy in the play sequence. The third toy was the target object which was unknown to E1 in both the newness and the givenness condition.

²² Because of the natural realization of the stimuli, microprosodic effects within the speech signal can not be excluded. Thus, another important reason for the reliability was to make sure that no uncontrollable microprosodic variations within the speech signal could have changed its perception.

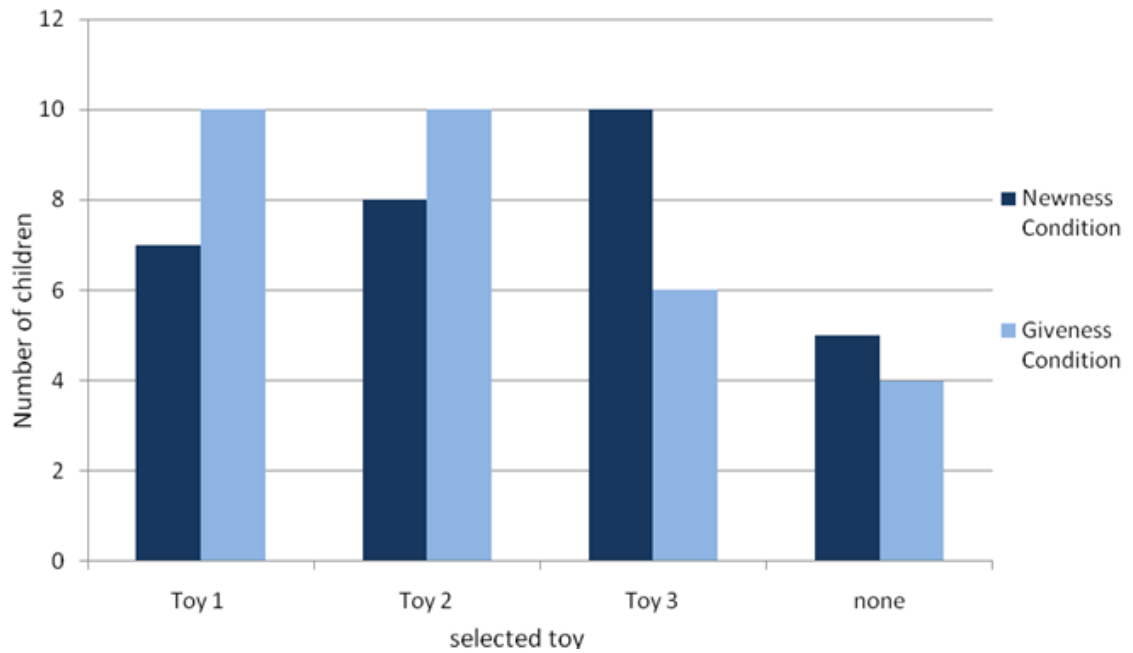


Figure 10: Results from this study. The diagram shows the number of children and the objects they chose in both the newness and the givenness condition.

I compared the number of children who chose the target object (Toy 3) with the target choices expected by chance in each of the conditions using the binomial procedure. Children in the newness condition selected the third, new-to-the-speaker object at chance level (10 out of 25, chance level: .33, $p=0.12$, one-tailed). If, however, children's choices in the givenness condition were compared with chance, I found that children handed over one of the "old" toys (object 1 or 2) more than would be expected by chance (20 out of 26, chance level: .67, $p=0.09$, one-tailed).

What these results show is that young children at the age of 20 months use information that is provided by intonation in order to find out what another person is referring to. This is especially interesting in the givenness Condition. Earlier research has mainly concentrated on children's understanding of another's knowledge regarding new and interesting objects. However, in this study, I found that children also understand what is old and already known for another person. Interestingly, the results for the newness condition did not show any significant preference for the object that the speaker did not previously see. One possibility which would explain these results is that children at this early stage simply need several cues in order to find out that the speaker is referring to a new object. As already mentioned, in earlier research the request in the experimental conditions presented a whole package of cues (e.g. lexical items, hand gestures, facial expressions, and intonation) in order to make the request clear. Children could use all of these cues, i.e. the whole "package" in order to find out what the person is referring to. Thus, it could be that children this age

need 'more' excitement behind a request, which has to consist of several supporting cues. What this means is that intonation alone seems not to be strong enough for children 20 month old children to understand that a person is referring to something that is new to him. This is consistent with the findings in the givenness condition. When a speaker is bored and disinterested in something (because he already knows it) he does not use excited cues. Thus, if a request is pronounced in a boring and uninterested way, children could have come to understand that this request refers to an old and known object. Additionally, it could be that children were confused about the use of the definite article 'den' in the request ("Gib mir mal **den** Flomer"). A definite article refers to something that is already established within a discourse and children could have assumed that the experimenter is referring to one of the old objects because of the use of this article (see Matthews, Theakston, Lieven & Tomasello, 2009)

The overall pattern of results suggests that children at the age of 20 month understand the difference between the different intonational realizations of a request. Depending on the speaker's previous knowledge, the children in this study understand that a typical givenness pitch accent L* refers to an object that is already known from the previous discourse. And, they understand that the use of a particular intonational contour has an intentional reason – the speaker means something by using that particular way of talking.

To summarize, in this study I could show that children understand that prosodic salience has a function within an utterance; it can mark the referential intention of a speaker. As already mentioned, even prelinguistic children attend to contextually new elements, and they interpret adults' linguistic and nonlinguistic referential expressions as referring to these new elements (e.g. Tomasello & Akhtar, 1995; Akhtar & Tomasello, 1996; Moll & Tomasello, 2007). However, more important in this study are the findings that children at 20 months of age understand that the intonational marker for given information in German indicates referentially old and shared information. 20 out of 26 children identified the referent marked as given correctly to the discourse. This means that young children are not only sensitive to what is new to another person, they rather understand what that person already knows. And, as a main finding, they can map the intonational realization to that knowledge.

The question that follows is whether children can use this strategy to acquire new words. Therefore, I did a second study in which I wanted to find out what role prosody plays within the word-learning process. To do so, I used a similar design as in the study presented in this Chapter and added a further cue. This new cue either supported intonation or conflicted with it.

6.2. Competition in Word Learning: Intonation vs. Mutual Exclusivity

6.2.1. Introduction

Children can, as we have seen in the previous study, use intonation among other social-pragmatic cues to infer certain aspects of the communicative intention of a speaker. This is an important source of information which, along with the nonlinguistic context, can elevate some interpretations about a word's meaning (Baldwin, 1995; Tomasello & Akhtar, 1995). However, as the results from the previous study suggest, there are multiple, sometimes redundant sources of information that children can use to interpret a novel term (Markman, 1992; Woodward & Markman, 1998). In some situations, these cues are not easy to interpret and can be uninformative or ambiguous. In addition to reference based on knowledge e.g., the state of newness, children can use another indirect cue in order to determine what a person is referring to. The "mutual exclusivity" constraint leads children to the assumption that each object has one and only one label (e.g. Markman & Wachtel, 1988; Merriman & Bowman, 1989, Diesendruck & Markson, 2001, Markman, Wasow & Hansen, 2003). Mutual exclusivity enables children to successfully infer the referents of novel terms, even when direct cues are missing. For example, in a situation in which a speaker does not point to or direct the child's attention to an object in any other way, the child cannot determine what object a novel word maps onto. Suppose, for example, a child sees two objects. One of these objects is familiar (e.g. 'dog'), while the other object is completely new to the child (e.g. 'stapler'). The child hears someone saying: "Can you hand me the stapler?" According to the mutual exclusivity assumption, a child should reject a second label for the dog-object and consequently infer that the word 'stapler' refers to the unfamiliar object (given that it is the only other object around). Thus, mutual exclusivity is an important instrument in order to find out the correct referent for a word. In order to further investigate the role of intonation in the word-learning process, I used the mutual exclusivity cue and put it either in contrast with intonation or used it as a support for intonation.

6.2.2. Data & Method

In this study, I wanted to find out what role intonation plays in the overall context of different cues. Additionally, I wanted to investigate if intonation is strong enough to overwrite mutual exclusivity, i.e. the fact, that every object has only one label and that new objects are automatically linked to a novel referent. The crucial difference as opposed to the study presented in the previous section was that the intonational cue was put in contrast to mutual exclusivity. Thus, whereas in the givenness Condition both the cue that is provided by mutual exclusivity and the speaker's intonation converged onto the same referent, namely a novel object which the speaker had previously seen, the mutual

exclusivity cue and the speaker's intonation contradicted each other in the newness Condition.

Participants

Subjects from Study 1 also participated in Study 2.

Materials, design, and procedure

The materials, design, and procedure were the same as in Study 1 that is, two experimental between-subjects conditions were created. The children's task was to identify the referent of a request. The only cues the children had was on the one hand the knowledge about different objects which the requesting person had and on the other hand the intonational form of a request. The word used in both conditions was a phonotactically correct disyllabic German pseudo-word ('Miemel') which was embedded in a typical and appropriate German request. The main difference between conditions was the kind of accent of the request: In the newness condition I used the typical marker for contextually new referents (H*), in the givenness condition the referent of the novel object was marked by the appropriate marker for given referents (L*). However, there was one crucial difference as opposed to the procedure of the study presented in the previous section. Instead of using three novel objects, which were all unfamiliar to the children, I used a familiar object as first toy (a shoe), an unfamiliar object as the second toy (a wooden ring) and a familiar object as the third toy (a house) (see Figure 11).



Figure 11: test objects used in this study. The pictures show a shoe (A), a wooden ring (B) and a house (C). Toy (A) & (C) were treated as known and familiar objects by children of this age, Toy (B) was treated as an unknown object.

Since the children from study 1 also participated in this study, the procedure followed that of study 1. The procedure was exactly the same as in the previous study that is, the two experimenters played with the child using the first two objects for about 40 sec. They showed the child how to manipulate the objects and commented on the joint action in a very general fashion, saying,

“Look at what you can do with this!” and “That’s nice!” None of the novel objects was labeled during play but pronouns were used (the German equivalent for ‘it’, ‘this’, or ‘that’). After the 40 second play-phase, E2 took the toy and placed it on a tray out of the sight of the children, saying, “I’ll put this here!” E1 left the room under a pretense before the third toy came out. He stood up, waved to the child and E2, saying: “Bye, Child! Bye E2!” After he had left, E2 advised the child that E1 was out of the room and that he could not see or hear them, but that they would play with another toy. She then brought out the third toy and repeated the same procedure as for toy 1 and 2. After they finished playing with the third toy, E2 took the tray with the toys on it and put it on the edge of the table. She then put an additional, empty tray opposite the child. Both of the trays were out of the child’s reach. E2 passed the objects from tray to tray in a counterbalanced order saying for each object: “I’ll put this here!” All utterances were realized with the same intonational pattern. In doing this, the child once again had the chance to watch all of the toys. E1 then came back and said: „Hello, I’m back“. He remained in front of the table at a distance of approximately 1 m. At that moment, E2 held the tray with the toys on it in front of the child, so that all objects were equidistant from the participant. E1 watched the toys for about 3 sec. Then he said: “Ah, Child, give me the *Mieme!*” He then approached the table and held out his hand to enforce his request. In order not to give the child any cue, he held his hand toward the middle of the tray at an appropriate distance, looking the child in the eyes. He repeated his request up to two times if necessary. Figure 7 (see above) summarizes the procedure.

Acoustic properties of the test material

In the newness condition, the intonational realization of the nonsense word ‘Miemel’ was marked with an H* with a preceding rise, high fundamental frequency, wide F0-excursions, expanded duration of utterance and pauses and a lower speed of articulation. In the givenness condition the intonational realization of the nonsense word ‘Miemel’ was marked by an L* pitch accent, characterized by lower fundamental frequency, narrower F0-excursions, shorter duration of utterance and pauses and a higher speed of articulation. The acoustic speech signal of the request was analyzed for the length of the utterance and the target word as well as the mean time at which the pitch accent reached its target within the word. Additionally, the mean frequency of the pitch accent was measured. Table 6 shows the analysis of the speech signal of the request.

Table 6: acoustic properties of the target utterance and the target word in both the newness and the givenness Condition

	mean length (utterance)	mean length (target word)	Mean time pitch accent (within target word)	Mean F0 (target word)	
				low	high
Newness Condition	2545,93 ms (132,07)	1184,30 ms (82,36)	293,47 ms (52,97)	150 Hz (52)	286 Hz (18)
Givenness Condition	1163,43 ms (69,51)	531,67 ms (44,64)	192,11 ms (61,52)	71 Hz (10)	

The intonational contour for the referent in the newness condition is characterized by a rise from about 150 Hertz to 286 Hertz (this corresponds to an average difference of 13,47 semitones), whereas the given request is marked by a flat contour with a low pitch accent at 71 Hertz. Even so, the pitch accent in the givenness Condition is realized earlier than in the newness condition²³. However, like in study 1, the distinction between the requests for a new respective given referent is realized by a greater F0-excursion, a different kind of F0-contour and pitch accents and by a different length of request.

Coding and reliability

The first experimenter did a live coding and judged which of the three objects the child handed over. Additionally, all test sessions were recorded, making it possible to do a control coding immediately after the session. To assess inter-rater reliability, a research assistant, who was unaware of the condition, coded 20 % (12 participants) of the final sample from the video material. The agreement between the two raters was 100%, for a Cohen's kappa of 1. In addition, 50% (15 new, 15 given requests) of the intonational realization of the request "Give me the *Mieme!*" was tested by a blind coder and was compared to the speaker's intention during the test-phase. Agreement between the two raters concerning the intonational intention was 100%, leading to a Cohen's kappa of 1.

²³ In some models on Intonation, the timing of a pitch peak has played an important role. For example, the Kiel Intonation Model (Kohler, 1991a) assumes that for the understanding of the paradigmatic dimension (e.g. the cognitive status of a target referent), the timing of the pitch peak (e.g. early, medial, late) is of essential importance.

6.2.3. Results & Discussion

In this study I added mutual exclusivity as a further cue. In the newness Condition, mutual exclusivity conflicted with newness-to-the-speaker and the speaker's intonation. In the givenness Condition, mutual exclusivity and the speaker's intonation converged to the same referent, namely a novel object which the speaker had previously seen (Toy 2). Figure 12 shows the number of children's object choices separately for the two conditions, with 'Toy 1', Toy 2' and 'Toy 3' referring to the temporal position of the toy in the play sequence.

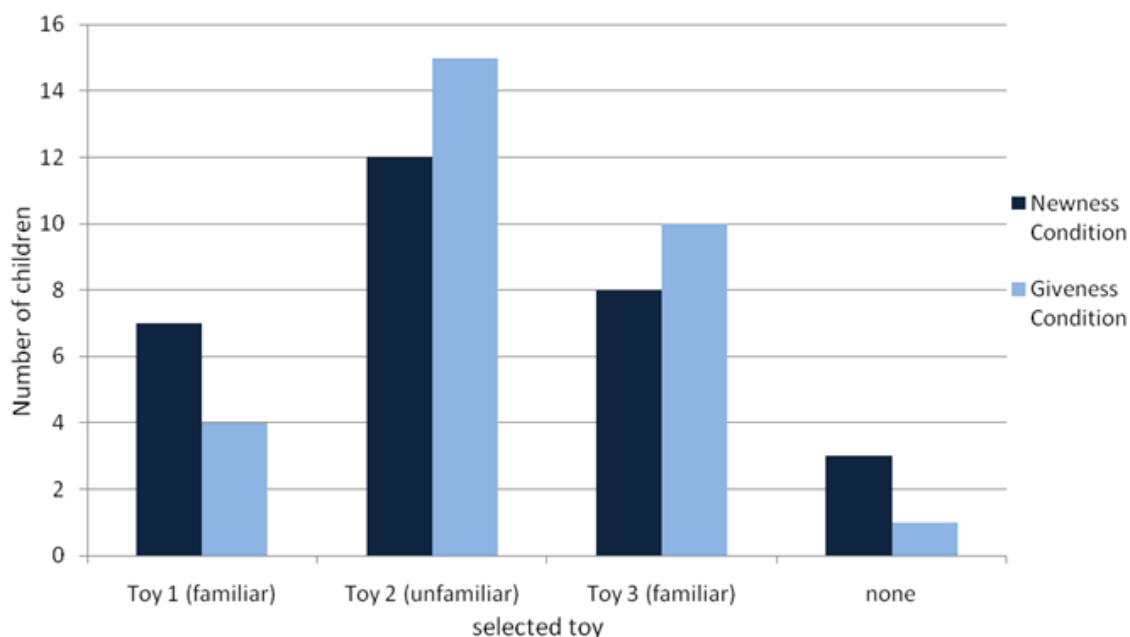


Figure 12: Results from this study. The diagram shows the number of children and the objects they chose in both the newness and the givenness condition.

I compared the observed number of children choosing the target object with chance using the binomial procedure. I found that children in the givenness condition choose the novel object that the speaker had previously seen (Toy 2) more than would be expected by chance (15 out of 29, chance level: .33, $p=0.01$). In the newness Condition, when mutual exclusivity conflicted with newness - intonation, children choose the "given" novel object (Toy 2) only marginally more than would be expected by chance (12 out of 27, chance level: .33, $p=0.07$). Comparison between conditions revealed that children's reliance on mutual exclusivity did not differ with intonation. In the givenness Condition 15 children relied on mutual exclusivity and 10 on newness and in the newness Condition 12 children relied on ME and 8 on newness ($\chi^2=0.82$, $p=0.365$).

What these results show is that children in the givenness condition chose in 15 out of 29 cases the second unfamiliar but known-to-the-speaker object. This was expected because both the intonational information as well as the information conveyed by the novel label point to that object. Interestingly, 10 out

of 29 children in this condition also chose the third, new-to-the-speaker, but familiar object. One explanation for this could be that the children just recognized that the requesting person had not seen that object before and that this intention was simply that he was automatically interested in that. If, however, the requesting person asked with an excited intonation, but an unfamiliar label for one of the three objects, children also relied on mutual exclusivity and chose the second object in 12 out of 27 cases. But, 8 out of 27 children in this condition relied on the intonational cue and chose the third, new-to-the-speaker object. Thus, it seems that for children of this age, they do not simply concentrate on one cue. As I concluded from the previous study, children seem to rely on several cues. And, as soon as some of these cues contradict each other, children of this age seem to be confused. This is also supported by the number of children who chose the first toy. 7 out of 27 children chose the toy which was presented first in the playing phase, although neither the mutual exclusivity nor the intonational cue pointed to that object. On the one hand, children could have had the problem that they knew the experimenter had not seen the third object, but that he was asking with a novel word (pointing to the second object) in newness intonation (pointing to the third object). This confusion could have led them to choose that object which was totally 'out of the game'. Additionally, the first object could have been the most salient one (because it came first in the playing phase).

To summarize, this study strengthens the suggestions from that study presented in Chapter 6.1., that is, children need several, supporting cues to lead them to an understanding about what another person is referring to. One, and only one cue, seems not to have the power to inform children of another's intention. According to intonation, the results from this study suggest that it is a very important cue that children use in order to acquire information about what another person is referring to. And, intonation seems to have the strength to pull children away from their strong reliance on mutual exclusivity, at least at this early stage in language acquisition.

6.3. General discussion

In the current studies, I found that children at the age of 20 months use different pitch accents in order to find out what another person is referring to. This was especially the case when the speaker used the typical givenness intonation and requested an object that he had already seen. However, even when the results for requesting a new-to-the-speaker object were not significant, the number of children who chose that object when requested with the appropriate intonation leads us to the conclusion that intonation is an important cue for young children in order to read the intention of a request. However, comparing the results with those from previous studies, it becomes clear that young children do need a combination of several cues, one of which is intonation.

Previous studies have shown that children are sensitive to discourse novelty. In order to understand that a speaker is referring to an object that he has

not seen before, the child has (1) to know that the speaker has not previously seen the object in this discourse context; and (2) to believe that an adult will name a novel object for a child when, in the discourse context, the adult and the child first jointly encounter the object. Thus, in previous studies, the task for the child was simply to identify the object that the speaker has not seen before. What is new in this study is that I could show that children also understand what object the requesting person had seen before and, more important, that the child can map the intonational form of that request to this experience. Thus, the child understands the intentional function behind the intonational form even when this goes against an expectation.

In the second part of this study, I added mutual exclusivity as a further cue, either supporting the intonational form or did not. Although most of the children (27 out of 56) chose in both conditions the second object and thus, relied on mutual exclusivity, 18 out of 56 children also chose the third, new-to-the-speaker object.

One could argue that the children in the newness condition reacted to the intonational form of the request. But, this argumentation is not sufficient enough to explain the behaviour of those children who chose the new-to-the-speaker object in the givenness condition. Overall, the results indicate that children were somewhat confused by the whole situation in which the cues contradicted each other. This supports the hypothesis that children, when acquiring language try to rely on several cues of which intonation provides a rich source of information, as I will show in this thesis.

When referents are new to a situation in some way, the speaker uses sentence accent to direct others' attention to this referent (see Chafe, 1994). Thus, if a mother says "Look, the boy has a nice DOGGIE", she probably wants her child to attend primarily to the dog. On the other hand, if she says "Look, the BOY has a nice doggie" she probably wants her child to attend primarily to the boy. In this situation, the child has to understand that the important part is accented and thus, more salient within the speech stream. The informational meaning conveyed by this behaviour is for example 'surprise' or 'agitation'. This shows that the speaker has a certain intention when marking information in a certain way. As the results demonstrate, children can understand the communicated surprise or newness based on the intonational form. However, vice versa, this also means that the child has to understand the mother's intention about the relative unimportance of the unstressed referents in the context. Even if the findings of Tomasello & Haberl (2003) and Akhtar et. al. (1996) could not be replicated, my studies show a tendency for the fact that children, when hearing an exaggerated and excited request, understand that the adult is referring to the object he has not seen before. However, the question remains as to why children in the newness condition are not as successful as in other studies. This could be due to the fact that, in order to understand the excitement behind a request, they need several supporting cues e.g., pointing and/or facial expressions. In the first study, children could only use one cue in order to find out what the other person was referring to. This is consistent with the findings in the givenness condition because a speaker who is 'bored' and disinterested in something, does not use excited cues. Instead, the request is uttered in an uninterested way and the child

seems to understand that this request refers to an old and known object. More generally, children understand that there are typical intonational patterns which are used in order to refer to the status of objects within a discourse.

To summarize, the results from these studies show that young children do already use intonation in order to interpret another's intentions. However, intonation on its own does not seem to be strong enough to do the job. Instead, it seems that children need a plethora of information in order to find out what another person is referring to.

7. The role of intonation in grammatical constructions

7.1. Resolving syntactic ambiguities

7.1.1 Introduction

The previous studies as presented in Chapter 6, showed that young children can use information that is conveyed by intonation in order to find out what another person is referring to. Consequently, the question arises whether the understanding about the intonational form as transmitter of a certain meaning continues with age. In the following chapters, I will present empirical evidence that deals with the question whether children use information that is conveyed by intonation in order to understand and to interpret more abstract grammatical constructions.

To acquire a language involves more than just the learning of words and grammatical rules. Children also have to learn how to interpret words and sentences by connecting them to the overall situation and the larger context. And, to become competent with language young children must master many different grammatical constructions: pairings between patterns of language use and their relatively complex communicative functions. A construction of particular importance in this process is the basic transitive construction, prototypically used to describe an agent acting on a patient. Children can use this kind of construction to describe the world around them e.g., various physical and psychological activities that people perform on objects. Thus, the basic transitive construction is typically produced in children's spontaneous speech early in language development (Tomasello 2003) and, developmentally, it is the earliest type of construction. But, before they can do this, they must learn and understand grammatical cues to determine the different roles of the two participants involved. Let us consider a novel transitive construction like the following example:

(11) 'The Flomer weefs the Miemel'

If one wants to understand and interpret such a construction with novel words (a situation children are exposed to every day), one not only needs to understand the meaning of the different words, but also certain rules of the particular language. A relatively easy task would be to understand a simple construction like 'The Flomer tamms' because there is only one acting participant involved (the 'flomer') who is performing an action ('tamming'). When a second participant gets involved, as in (11), the situation gets more complicated because one has to understand who is doing what to whom. Interestingly, in most languages the listener has multiple, sometimes redundant cues to acquire these rules, e.g. word order, case marking, or animacy – and, children from different language groups differ in their reliance on these cues from an early age. For example, if we take an English sentence like 'She eats the apple', a speaker of English can use several cues which can be reliably trusted in order to understand who is doing what to whom in that example. It is more or less easy to identify 'she' as the subject and thus, as the agent of the sentence, because (a) it is said before rather than after the verb (word order) (b) it is the subject pronoun and not the object pronoun "her" (case marking), (c) it agrees in number with the verb (verb agreement) and (d) it is commonly assumed that animate beings, here realized as the female pronoun, are more likely to act on inanimate things, than the other way around (animacy). An English learning child could use one or all of these cues to determine the participant roles in the acquisition process of transitive sentence like „The Flomer weefs the Miemel“ and she can use these cues to learn and to understand the grammatical rules of the particular language that are needed to understand different participant roles. However, depending on the language environment in which a child grows up, the cues that she can rely on will differ. One framework to consider how, when and in which order children acquire different cues in different languages is offered by the Competition Model of Bates and McWhinney (1987, 1989). The Competition Model is clearly a Usage-Based model in the sense that it ties the development of children's grammar to particular features of the input – the relative weights of individual cues. It is based on the psychological mechanisms that bring together different cues with their validity or information value. Cue validity is the product of two components: cue availability (how often is the current cue available over the total amount of cases) and cue reliability (how often does the current cue lead to the correct conclusion). Cue validity differs with language, because different languages rely on different cues. Most of the studies done within the framework of the Competition Model concentrate on this, i.e. how are participant roles marked linguistically in various languages and how do children learn and use these cues in sentence processing. In the typical Competition Model experiment, subjects are asked to choose the agent in sentences in which two or more cues conflict with each other. For example, in the following examples, word order is in direct conflict with agreement (12) and animacy (13):

(12) The girl chase the boys.

(13) The ball pushes the boy.

In both examples, subjects should choose the first NP ('the girl' or 'the ball') as agent if they followed word order as a cue to agent-patient relations. If, however, they followed agreement, they should pick the second NP ('the boys') in (12) as agent, and if they followed animacy, they should pick the second NP ('the boy') in (13) as agent. MacWhinney, Bates, and Kliegl (1984) compared English, Italian, and German and found that English-speaking adults always rely on word order to determine this kind of agent-patient relations. German-speaking adults also take agreement and animacy into account, and Italians most strongly rely on agreement. For our examples above, this means that English speaking adults would always pick the first NP as agent in examples (12) and (13), whereas Germans would pick the second NPs in both examples and Italians would pick the second NP in example (12) and presumably also in example (13). These experimental findings can be explained by the fact that English has very strict SVO word order. For example, the vast majority of English sentences have a fixed SV(O) word order and thus, a fixed order of agent and patient. Due to the fact that agents almost always precede patients, English-speaking children and adults consistently interpret the first NP in an utterance as agent and the second NP as patient. Additionally, agents are usually animate, whereas patients are often inanimate. This detail becomes more crucial when one considers how sentences in languages with variable word order such as Italian and Chinese are processed and interpreted. These languages are often determined by pragmatic factors. Thus, instead of paying attention to word order, Chinese- and Italian-speaking children and adults decide who is agent and who is patient on the basis of animacy (Bates, MacWhinney et al., 1984; Chan, Lieven & Tomasello, 2009). In a comprehension task in which American and Italian children between the ages of 2,5 and 5,5 were required to predict the role of agents and patients, Bates et al. (1984) compared sentence interpretation strategies from these two language groups. Their findings show that children from an early age use the most reliable cue for agent-patient relations of their mother-tongue – word order for English learning and animacy for Italian-learning children.

However, how these cues interact either during online processing or in the process of development is still an open question. One possibility is that children start by relying on only the most 'valid' cue for their language, only subsequently developing sensitivity to less 'valid' cues as they build up their strength. An alternative is that children may initially rely on a 'sentence schema' (cf. Chapter 3.2.) in which all, or most, of the cues are present and only subsequently abstract the relative value of each cue. Thus in the Dittmar, Abbot-Smith, Lieven and Tomasello (2008) study, discussed in more detail below, the youngest children were only able to correctly identify the subject of the sentence when it was

marked by both case and SVO word order, reflecting the 'coalitions-as-prototypes' suggestion of the Bates and MacWhinney (1987) model. This would fit with evidence that children start by learning form-meaning patterns in which child-identified meanings are connected to 'schemas' which are only partially analyzed into the components of adult grammar (for instance the 'whole word' approach in phonology, Vihman & Croft, 2007; and 'schema' learning in syntax, Tomasello, 2003; Dąbrowska & Lieven, 2005; Bannard, Lieven & Tomasello, 2009). By the time children are five – the age of the children in the Dittmar et al. study – one would expect them to have gone some way towards identifying these cues and their particular role in the construction. In addition, morphological (e.g. case-marking), intonational (e.g. focus) and syntactic constructions (e.g. 'grammatical subject') are also being gradually abstracted on the basis of form and function relationships between constructions.

However if children are indeed initially learning a schematic version of constructions then it is highly likely that, in real life, prosody is an essential component because constructions have a characteristic prosody (Taylor, 2002). In terms of the grammatical use of prosody, some researchers have found that it has little or no effect on children's interpretation of structurally ambiguous sentences (e.g. Vogel and Raimy, 2002, Choi and Mazuka 2003; but see Snedeker and Yuan, 2008, for more positive results using both action and looking time measures). But, as already mentioned in Chapter 4.2., Arnold (2008) recently found that 4- and 5-year-old children use the presence or absence of sentence accent to guide their interpretation of the degree to which noun phrases are given by the discourse context. A number of studies have shown that adult listeners use prosodic cues reliably to resolve syntactic ambiguities (Schafer, Speer, Warren and White, 2000) and to find phrasal boundaries (e.g., Carlson, Frazier and Clifton, 2009; see Speer, Warren and Schafer, 2003, for a review). Very few studies, however, have focused on the use of intonation to assign basic participant roles, such as the agent and patient in transitive sentences. In the framework of the Competition Model, Bates et al. (1984) found that 3.5-year-old Italian children used accentuation as a cue, but only in interaction with non-canonical word order (and the effect went away with older children). A language where intonation might be even more important for interpreting transitive sentences is German. While most transitive sentences in German have agent-patient word order (with the main verb in either verb-second or verb-final position), word order can be variable, with the patient sometimes coming first. The inviolable cue for agent-patient relations is thus case marking, which occurs on the determiner. The problem is that the case system has been prone to much syncretism, and so sometimes case marking is ambiguous. The following examples illustrate the situation.

- (14) Der Löwe VERB den Hund. [word order and case both usable]
The-masc-nom lion VERB the-masc-acc dog.
- (15) Die Katze VERB die Ziege. [case marking ambiguous]
The-fem-nom/acc cat VERB the-fem-nom/acc goat.
- (16) Den Hund VERB der Löwe. [word order & case marking conflict]
The-masc-acc dog VERB the-masc-nom lion. [lion is agent!]

In (14), the prototypical example, word order and case marking both indicate the first noun phrase as the agent. In (15), case marking is ambiguous and thus it is unclear whether the first noun phrase is the patient and the second noun phrase is the agent or vice versa. In this case, word order is typically used (i.e. again identifying the first noun phrase as the agent). In (16) - a so-called patient-first sentence - case marking and word order conflict and, due to the nature of German grammar, case marking prevails (and the preverbal noun phrase is the patient). A construction like this where the first noun phrase is the patient is much less common in German, and it therefore typically occurs with a prominent accent on the first noun phrase.

Weber, Grice & Crocker (2006) examined whether prosody, beyond other cues such as case marking, can manipulate the interpretation of word-order ambiguities for adult listeners. They tested German adults using an eye tracking paradigm and presented sentences with case-ambiguous first NPs and unambiguous second NPs, e.g

L*+H

H*

- (17) „Die Katze (ambiguous) **jagt** womöglich den Vogel (+accusative)“

“The cat is possibly chasing the bird.”

L+H*

- (18) „**Die Katze** (ambiguous) jagt womöglich der Hund (+nominative).“

“The cat is possibly chased by the dog.”

In order to examine the influence of prosody on listeners' interpretation of participant roles, the agent-first utterance in (17) was intonationally realized by a low pitch accent (L*+H) on the first NP and H* on the verb, typically used for canonical agent-first sentences. For the Patient-first utterance in (18), the realization of the first NP was marked by a rising pitch accent (L+H*), expected to indicate non-canonical patient-first sentences. The results show that participants, immediately upon hearing the first noun phrase, fixated on the agent of the action (in a picture depicted by the sentence) when the nuclear accent (sentence stress) was on the verb, typically used for canonical agent-first sentences, as in (17). In contrast, when the realization of the nuclear accent was on the first NP, typically indicating non-canonical patient-first sentences, participants interpreted the ambiguously case-marked, first noun phrase as the patient, as in (18). These results show that adult-listeners do use intonational information in the interpretation of ambiguous SVO and OVS sentences when no clear morphological information is available. Before the onset of the second NP, the patient was fixated upon more often than the agent when the intonational pattern already indicated the first NP as the agent, but not when intonation pointed to NP1 as the patient. Participants attended to and used intonational information to guide their comprehension of such sentences. Thus, the interpretation of word-order ambiguities was modulated by prosody and this was integrated rapidly enough to affect listeners' interpretation of grammatical function and assignment of participant roles before case information became available to clarify the ambiguity.

Dittmar et al. (2008) investigated young German children's comprehension of transitive sentences (containing nonsense verbs) that had various combinations of word order and case marking cues (see examples (14) – (16)). They found that children as young as 2.5 years of age had a strong word order bias. They only correctly interpreted transitive sentences in which both word order and case marking indicated the first noun phrase as the agent. But when word order and case marking conflicted, as in (16) above, only the 7 year-olds behaved like adults by relying on case marking over word order. That is to say, the 2-year-olds and 5-year-olds most often interpreted the agent in sentences such as 'Den Hund verb der Löwe' as being the first noun phrase, whereas adults chose the second noun phrase almost 100% of the time. The problem, however, is that in this study all of the sentences were produced for the children with very similar prosody for all conditions. But, patient-first sentences are not felicitous if they do not have the typical OVS-marked intonational pattern. It is therefore possible that young children are capable of understanding patient-first transitive sentences but only when the natural intonational pattern that they hear in their everyday environment is present (as it was for the German adults in the Weber et al., 2006, experiment).

In the current study, therefore, I used a paradigm very similar to that of Dittmar et al. (2008) but systematically varied prosodic cues. In two studies, I presented five-year-old German children with transitive sentences involving nonsense verbs (so that they could not use verb-specific information to interpret the sentences). Both studies employed a 2x2 design. Sentences either had

ambiguous case marking or else they were marked by case on the determiner as patient-first sentences (the kind that children systematically misinterpreted in the Dittmar et al. study). Crossed with this variable, I either provided or failed to provide a rising L+H* pitch accent on the first noun phrase (of the type successfully used by German adults in the Weber et al. 2006 study). The question was thus whether children would use pitch accent on the first noun phrase in an adult-like manner to interpret transitive sentences and move away from their strong word order bias – both when case marking indicated that the first noun phrase was the patient and also when case marking was ambiguous so that accentuation, in a sense, competed with word order. The prediction was that 5-year-old children should be able to use the cue provided by intonation, and so show more skill with non-canonical, patient-first transitive sentences than children in the Dittmar et al. study. If so, it would be the first study to my knowledge in which young children systematically use prosodic information, intonation in particular, as a grammatical cue to assign basic participant roles during sentence interpretation.

7.1.2. Data & Method

Following Dittmar et al. (2008), in the first study, a video-pointing task was used to evaluate young German children's tendency to interpret transitive sentences on the basis of word order and case marking. I presented these sentences as either clearly case-marked (e.g. 'Den (+accusative) Hund wieft der (+nominative) Hase') or ambiguous (e.g. 'Die (+nominative / accusative) Katze wieft die (+nominative / accusative) Kuh'). What was new in the study was that I either did or did not provide a prosodic cue that indicates a patient-first interpretation for adults (Weber et al., 2006). To make sure that the prosodic cue was given appropriately and consistently, all sentences were computerized and manipulated regarding their intonation. The prerecorded stimuli were presented to children over a hidden audio speaker.

Participants

Sixteen monolingual German children with an average age of 4;10 years (range 4;5 – 5;3; 8 boys and 8 girls) were included in the study. An additional 2 children were tested but excluded from the study due to disinterest in the video clips (1) or experimenter error (1). Children were recruited from a database of parents who had volunteered to participate in psychological studies. They came from diverse socio-economic backgrounds. All children were tested in nursery schools in a medium-sized German city. As a control group, I tested 10 adults with the same procedure.

Materials and design

All novel verbs referred to prototypical causative transitive actions, involving direct contact between a volitional agent and an affected patient.

Actions were reversible and involved either a caused change-of-state or a change-of-location. The four novel verbs 'wiefen', 'tammen', 'baffen' and 'mommeln' were used to describe four novel transitive actions that were performed with four novel apparatuses. 'Wiefen' was used to refer to an animal rocking another animal, which was standing on an apparatus resembling a rocking-chair, by pushing it with its head. 'Tammen' referred to an animal pushing down on another animal by jumping on its back so that the platform it was standing on, with a spring underneath, sank. 'Baffen' was used to refer to an animal spinning around another animal that was standing on a disk. The fourth novel verb 'mommeln' referred to an animal jumping on a platform in order to catapult an animal standing on the other side of this catapult. (For test sentences and animal pairing see Appendix A). The agents and patients of a particular event were pairs of animals with the same grammatical gender. Exactly which gender depended on the condition. All children heard the same test sentences in four conditions: In Condition 1, the Case Marking / Contrastive Intonation condition, they heard the novel verbs within an argument structure in which the patient was the first noun phrase and was case marked with the accusative, and the agent was the second noun phrase and was case marked with the nominative; for example, 'Den (+accusative) Hund wieft der (+nominative) Elefant.' – 'The (+accusative) dog is weefing the (+nominative) elephant.' The intonational realization of the utterances was characterized by a strong pitch accent on the first noun phrase. In Condition 2, the Case Marking / Neutral intonation condition, children heard a sentence structure with the same grammatical markings as in Condition 1, but here, the construction was completely deaccented.

In the No Case Marking / Contrastive Intonation Condition, the German case marking was ambiguous (because only those animals were used that take the German feminine or neuter gender, which does not decline in the nominative or accusative case, e.g. 'Die Katze wieft die Ziege.' - 'The cat is weefing the goat.') and thus it was unclear whether the patient was the first noun phrase and the agent was the second noun phrase or vice versa. But, as in Condition 1 intonation was characterized by a strong, contrastive L+H* accent on the first noun phrase, which indicates NP1 as the patient. Accordingly, in the No Case Marking / Neutral Intonation structure, the children heard a sentence structure with the same grammatical markings, but with monotonised intonation. Each of the four conditions was tested with each of the four novel verbs; therefore each child heard 16 test sentences (see Table 7).

Table 7: Examples of the four test conditions containing the four novel transitive actions. That referent that was treated as agent is printed in bold.

Intonation Case Marking	+ strong contrastive Pitch Accent on NP1	- flat intonation
+	"Den (+accusative) Papagei wieft der (+nominative) Elefant" "The (+accusative) parrot is weefing the (+nominative) elephant"	"Den (+accusative) Hund bafft der (+nominative) Löwe" "The (+accusative) dog is baffing the (+nominative) lion"
-	"Das (+ambiguous) Zebra tammt das (+ambiguous) Eichhörnchen" "The (+ambiguous) zebra is tammimg the (+ambiguous) squirrel"	"Die (+ambiguous) Katze mommelt die (+ambiguous) Ziege" "The (+ambiguous) cat is mommeling the (+ambiguous) goat"

I tested each child with four different novel verbs in transitive sentence structures using a video pointing task. During the session, the children sat in front of a 23" TFT-screen (1920*1200 Pixel, aspect ratio 16:10). In the test trials, the child saw two film scenes on the computer screen, each starting simultaneously and lasting 6 s, followed by a still image of the clips. Both of these scenes involved animals enacting the same causative event and differed only in that the agent and patient roles were reversed. All children received alternating test sentences with the four different conditions and all four novel verbs were tested in one session. The order of the conditions and the novel verbs were counterbalanced in a 4*4 Latin square. The target screen order was counterbalanced so that the patient-first scene was presented on each side (left [L] or right [R]) in eight out of 16 trials for each child (e.g., for the pairing 'dog weef lion' and 'lion weef dog', half of the children saw the patient-first scene on the right initially and the other half saw it on the left, depending on counterbalance order). A particular side was never the correct choice for the patient-first scene more than twice in a row. No child experienced a test session in which the patient-first scene alternated regularly (e.g., LRLRLRLRL). The direction of the action was also counterbalanced (e.g. in the pairing 'dog weef lion' and 'lion weef dog' half of the children saw the agent performing the action from the left side of the scene towards the right side, and for the other half they

saw the reverse). In order not to take any cues from the experimenter, the test trial was conducted with a talking puppet. All auditory stimuli were prerecorded and uttered by the puppet.

Acoustic properties of the test material

The intonational realization of the utterances in Conditions 1 and 3 was characterized by a strong, rising L+H* pitch accent on the first nominal phrase (see Figure 13). Subsequently, the intonational realization of the utterances in Conditions 2 and 4 was characterized by a flat and monotonized intonational contour throughout the whole utterances (see Figure 14).

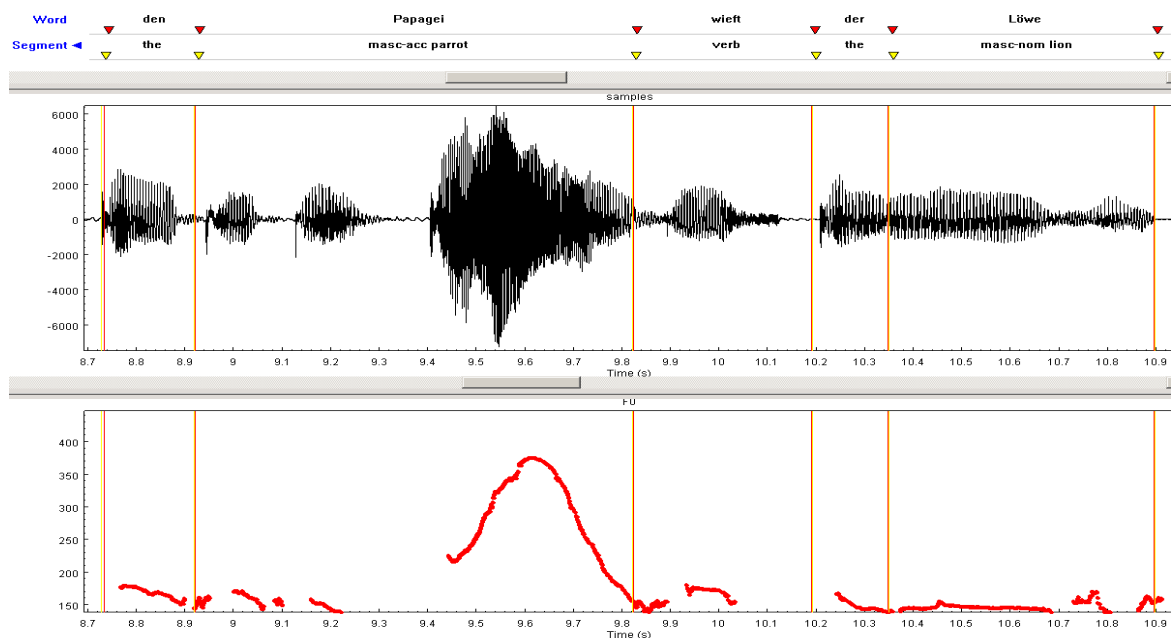


Figure 13: Example of the intonation of the target utterance in the Contrastive Intonation condition. The contour bears a L+H* pitch accent on the first Nominal Phrase.

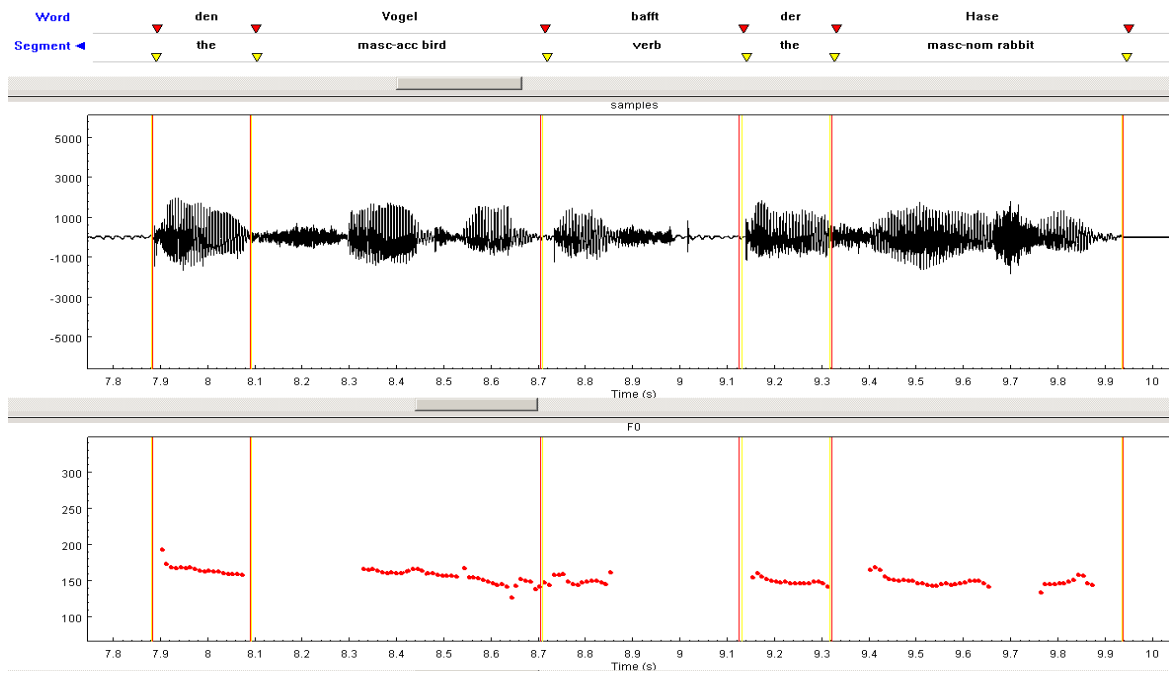


Figure 14: Example of the monotonised intonation of the target utterance in the Neutral intonation condition.

All stimuli were recorded by a female native speaker. She was asked to utter the sentences with as much emphasis as possible in the Contrastive Intonation conditions or as flat as possible in the Neutral Intonation conditions. If necessary, the recordings were later edited and manipulated by a speech analyst and ToBI-expert. He ensured that the stimuli were as natural as possible. An analysis about the acoustic properties of the test stimuli is shown in Table 8.

Table 8: acoustic properties of the test stimuli. The table shows the mean minimum and maximum fundamental frequency (F0) and the pitch range in semitones (st) of NP1 and the whole utterance plus the standard deviation in parentheses.

	NP1			utterance		
	minimum F0 in Hertz	maximum F0 in Hertz	pitch range in st	minimum F0 in Hertz	maximum F0 in Hertz	pitch range in st
Contrastive Intonation	131,53 (38,1)	384,25 (39,5)	19,27 (6,5)	105,85 (34,2)	384,26 (39,5)	23,08 (6,2)
Neutral Intonation	150,19 (7,3)	187,26 (20,2)	3,7 (1,14)	133,81 (26,7)	202,14 (25,1)	7,44 (4,0)

Procedure

The test session lasted for approximately 15 minutes. The computer monitor was positioned on the table approximately 50cm in front of the child. All sessions were videotaped with a camera centered behind the child, recording the child's pointing behaviour. The experimenter never looked at the screen during the test trials but sat behind the screen pretending to read.

Pointing practice training: To teach the children that the aim of the task was to point to one of two pictures on the computer screen, a very easy warm-up task with two pictures depicting objects was used; for example, 'cheese' and 'bacon,' appeared on the screen simultaneously. The children were then asked by the experimenter to point to one of the two objects (e.g., '*Zeig mir das Bild: Das ist der Käse.*' – 'Show me the picture: That's the cheese'). This task was repeated 10 times with different pictures and all children solved it perfectly.

Word learning training: Each of the novel verbs and the corresponding actions were presented to each child through a live performance given by the experimenter. To show and teach the different functions of the novel apparatuses, and thus the novel verbs, the experimenter performed the novel actions using animals whose labels take the German feminine gender and are ambiguous in the nominative or accusative case (e.g., '*Ziege*' – 'goat' and '*Ente*' – 'duck'). Each of the four novel verbs used in the test were randomly presented one after another by the experimenter in a variety of argument structures: in the citation form with no arguments (e.g., '*Das heißt wiefen.*' – 'That's called weefing'), as well as in a transitive argument structure with two feminine

pronouns (which are identical for subject and object position in German) in three different tenses ('*Sie wird sie wiefen.*' – 'She is going to weef her'; '*Sie wieft sie.*' – 'She is weefing her'; '*Sie hat sie gewieft.*' – 'She weefed her'). The child was asked to repeat the verb using a prescribed question format (e.g., '*Kannst du das sagen: wiefen?*' – 'Can you say that: weefing?') while the experimenter performed the action.

Film familiarization trial: Following the word learning training, the puppet declared that she has designed special clips which she wants to show the child and the experimenter; the child always agreed to see them. The child then received a familiarization trial for each verb in which he or she watched one film scene on just one half of the screen, involving two animals, with German feminine or neutral gender, acting out the novel verbs. At the same time, the puppet described the scene in a scripted manner; for example, '*Guck mal, das heißt wiefen.*' – 'Look, that's called weefing.'; all the while the other half of the screen remained blank. The side of the screen where the children saw the first picture (left or right), the acting direction, as well as the order of the novel verbs, was counterbalanced across and within subjects. At the end of each scene, the experimenter pointed to each animal and asked the child '*Wer ist das?*' - 'Who's that?' The majority of the children had no problem in spontaneously naming the participating animals. If a child did not name one of the animals, the experimenter told the child the name and asked him or her to repeat it, which nearly all of the children then did.

Test trial: The puppet then told the child and the experimenter that she had even more films that she would like to show. The experimenter then said that unfortunately he needed to read something and had no time to watch these clips with the child and puppet. He then sat behind the screen, and ran the computer program. Shortly afterwards, a red dot focused the child's attention on the center of the computer screen.

The test trial then began and the child watched two scenes simultaneously (see Figure 15 for an example display), which were accompanied by a prerecorded linguistic stimulus, explaining who was present in the clips and what they were doing; for example: '*Guck mal, das Schwein und das Pferd. Das heisst wiefen!*' – 'Look, the pig and the horse. That's called weefing!'.

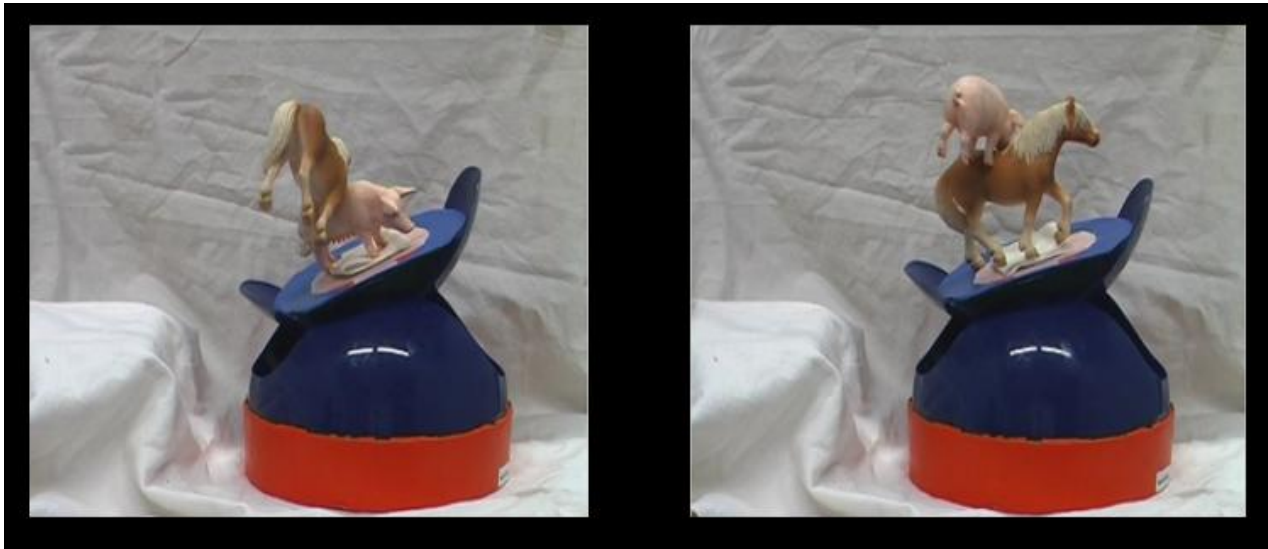


Figure 15: Example display about the material used in the test trial. In the left scene the horse is 'weefing' the pig, in the right scene the pig is 'weefing' the horse.

After the videos had stopped, the prerecorded voice of the puppet asked the child to point to the correct (still) picture by asking the target sentence according to one of the four conditions; for example, *'Zeig mir das Bild: Das Schwein wieft das Pferd!'* – 'Show me the picture: The (+ambiguous) pig weefs the (+ambiguous) horse!' If the child did not point to one of the two film scenes, the puppet repeated the question a second time; however, she never asked the child to point again once she/he had already done so. Once the child had pointed to one of the two pictures, the next test trial began, preceded once more by the red dot.

Coding and Reliability

For every test trial, I coded responses for whether participants pointed to the picture in which the post-verbal, second noun in the sentence was the agent. This was, of course, correct in the Case marking conditions, but either picture choice was possible in the No Case marking conditions. The question of interest is whether the addition of intonation would influence the children's choices. If a child did not choose either scene (= 2 trials), I coded those trials as 'wrong' (an alternative analysis in which these cases were excluded had no effect on the pattern of the results). All children were coded by the first experimenter, and an additional coder coded 25% of all trials for testing reliability (= complete session of four randomly selected children). This revealed a perfect agreement with the first rater (Cohen's Kappa = 1.0).

7.1.3. Results and Discussion

Children

I tested for the proportion of times the NP occurring after the verb was identified as the agent of the action out of four. The data were analyzed using a 2 (Intonation) x 2 (Case Marking) repeated measures analysis of variance (ANOVA)²⁴. There were main effects for both Intonation, $F(1,15)=4.88.4$, $p=.043$ and Case Marking $F(1, 15)=42.8$, $p< .001$, but there was no significant Intonation x Case Marking interaction, $F(1,15) = 3.608$, ($p=0.061$) (see *Figure 16*).

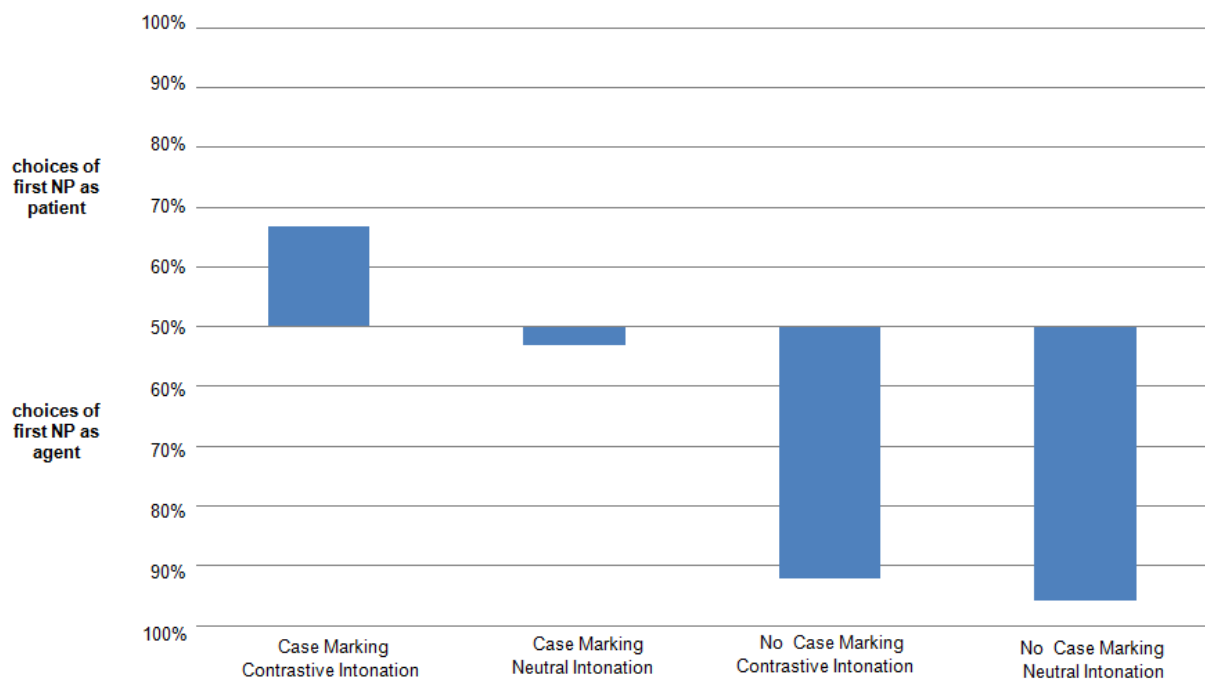


Figure 16: Results of the study in the four conditions. The diagram shows percentages of judging NP1 as either patient or agent as compared with chance, 50 %.

Because the chance level for the dependent variable was always 50%, I also investigated in which conditions the children were above chance in choosing the first noun as patient. The results show that the children were only above chance in the Case Marking / Contrastive Intonation condition (Condition 1; one sample t-test: $t(15) = 2,2$, $p=0.044$). In contrast, in the Case Marking / Neutral Intonation, the children were approximately at chance level (Condition 2; $t(15) = -.355$, $p= 0.728$) and in the No Case Marking / Contrastive Intonation (Condition 3)

²⁴ Additionally, the data has been analyzed using a General Linear Mixed Model. These results revealed the same overall pattern of results, i.e. significance values of interactions and main effects.

as well as in the Case Marking / Neutral Intonation Condition (Condition 4), children were below chance (both $t(15) < -14$, both $p > 0.01$), i.e. they were significantly more likely to choose the first noun as agent.

A comparison between the two conditions Case Marking / Contrastive Intonation and Case Marking / Neutral Intonation showed that children were significantly better in judging participant roles when intonation was available (paired-sample t-test: $t(15)=2.36$, 0.032). Choices in the two conditions No Case Marking / Contrastive Intonation and No Case Marking / Neutral Intonation were not significantly different ($t(15)=0.368$, $p=0.718$).

Adult - control group

For the adult – control group, I found main effects for Case Marking, $F(1,9)=50.08$, $p < .001$, but not for Intonation and no significant interaction between the two (see

Figure 17).

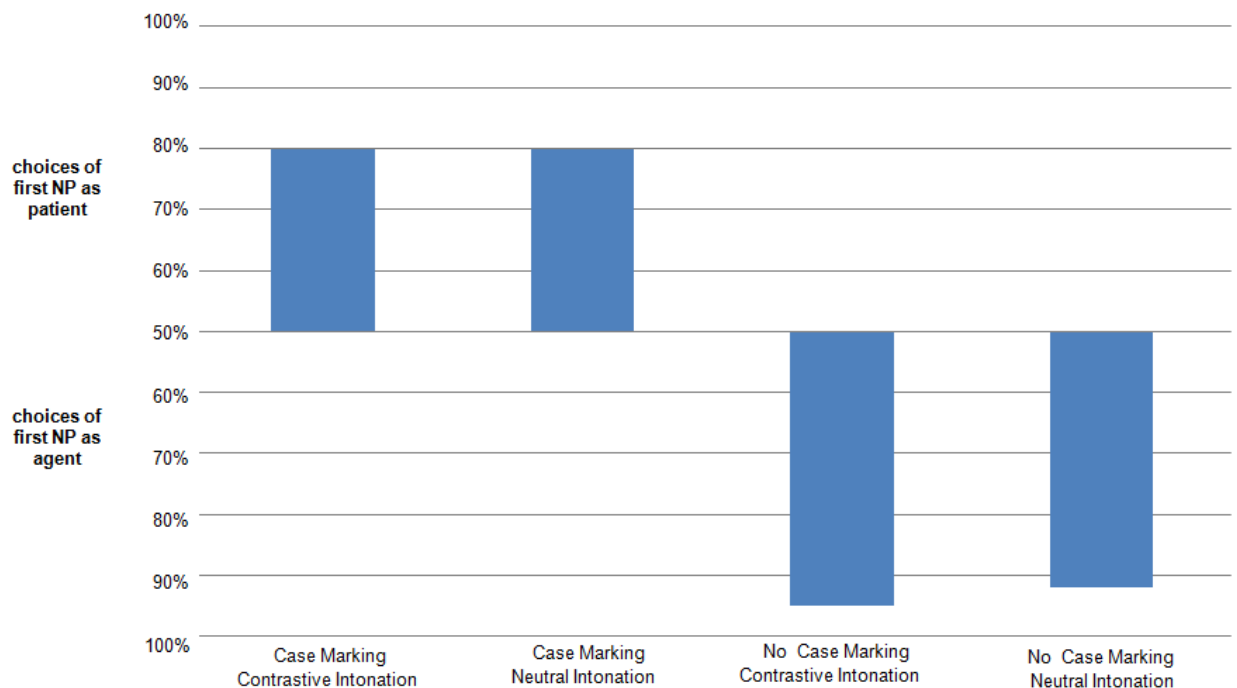


Figure 17: The results of study 1 for adults in the four conditions. The diagram shows percentages of judging NP1 as either patient or agent as compared with chance, 50%.

For a further analysis, I compared the results from the children with those of the adults. The data were analyzed using a three-way mixed analysis of variance (ANOVA) with two repeated factors (Case Marking and Intonation) and one between-subjects factor (Age). There were main effects for Case Marking, F

(1,24) = 96,72, $p < 0.01$, but not for Intonation, $F(1,24) = 3.12$, $p = 0.09$. There was a marginally significant interaction between Case Marking and Age, $F(1,24) = 4.49$, $p = 0.045$, but no significant difference between Intonation and Age ($F(1,24) = 1.9$, $p = 0.180$), between Case Marking and Intonation ($F(1,24) = 1.48$, $p = 0.235$) or between a three-way-interaction ($F(1,24) = 2.66$, $p = 0.115$). A comparison between conditions of the two studies only revealed a significant difference between children and adults in the Case Marking / Neutral Intonation conditions ($t(9) = -3.35$, $p = 0.008$).

What these results show is that the children are using case marking when it is available and word order when it is not, to interpret the roles of the NP's in transitive sentences. Thus, children moved strongly away from choosing NP1 as the agent when case marking indicated this as the patient. The two conditions without case marking show that intonation by itself is not sufficient for these young children to identify a transitive construction combined with the appropriate OVS-intonation pattern. This is consistent with the findings discussed in Chapter 6. They instead rely heavily on the word order cue, choosing therefore the first noun as the agent. Comparison of the two conditions with case marking however, suggests that the intonation and case marking together provide a stronger cue than case marking alone. This was not the case with adults who could use case marking alone to select NP1 as the patient. This shows that children can use intonation in order to glean extra information when it is used redundantly with other cues. This finding is broadly consistent with the findings of Dittmar et al. (2008) that German children best comprehend transitive sentences with multiple, redundant cues. In their study the two cues that reinforced one another were word order and case marking, and in the current study they were case marking and intonation.

7.2. The role of context & intonation in resolving syntactic ambiguities

7.2.1. Introduction

The test sentences from the study presented in the previous Chapter were presented to children outside of any meaningful discourse context. If intonational highlighting serves in many cases to contrast the stressed item with something in the previous discourse, then one could argue that presenting sentences in isolation does not provide children with a natural interpretive context and is, in fact, contrary to the principles of a Usage-Based approach. Indeed, in the adult literature, it has been argued on several occasions that experimenters should present intonationally contrastive sentences in more natural discourse contexts (e.g., Albritton, McKoon and Ratcliff, 1996). In the second study, therefore, I used the same linguistic materials and same basic method as in Study 1, with one crucial difference. Each test sentence was preceded by a

discourse context in which a speaker described a scene incorrectly by misidentifying the patient using a normal, agent-first transitive sentence (e.g., "The dog is weefing the frog", when in fact he is weefing the lion). The test sentence was then a patient-first transitive sentence, uttered as a correction, with an accent on the patient (in very loose translation, "No, it is the LION that's getting weefed."). This is arguably something close to the "natural home" of patient-first transitive sentences in everyday German discourse, and should give young children a better opportunity to show even more skills at using intonation to interpret patient-first transitive sentences.

7.2.2. Data & Method

Participants

Sixteen monolingual German children with an average age of 4;10 years (range 4;6 – 5;3; 10 boys and 6 girls) were included in the study. Children were recruited from a database of parents who had volunteered to participate in psychological studies. They came from diverse socio-economic backgrounds. All children were tested in nursery schools in a medium-sized German city.

Materials and design

Materials and design were the same as in Study 1 with the exception that the instructions for the test trials did not come from just one puppet, but instead were communicated in a conversation between two puppets. Whereas one of the puppets was the same character as in study 1 (P1), the other puppet (P2) was introduced as an unreliable character because he was too young to know the names of the animals or not able to remember the novel verbs. Instead, he said everything wrongly and was therefore corrected by P1. Thus, the target instruction in the form of the transitive OVS utterance (using the same stimuli as in study 1) was embedded in a contrastive context.

All children heard the same test sentences (see Appendix B) in a transitive OVS structure. The same four novel verbs were used in the same four conditions as in study 1: Case Marking / Contrastive Intonation, Case Marking / Neutral intonation, No Case Marking / Contrastive Intonation, No Case Marking / Neutral Intonation. Before the child heard the target sentence, P2 uttered a transitive SVO sentence, in which the patient was always wrong as in (19). P2 was then corrected by P1 using an utterance of the target sentence in transitive OVS structure, as in (20).

(19) Der Löwe verb den Frosch!

The-masc-nom lion verb the-masc-acc frog!

The lion verb the frog.

(20) Nicht den Frosch verb der Löwe, sondern den Hund verb der Löwe!²⁵

Not the-masc-acc frog verb the-masc-nom lion, but the-masc-acc dog verb the-masc-nom lion!

It's not the frog that's verb the lion, it's the dog that's verb the lion!

An example of the first part of the correcting utterance as in sentence (20) above can be seen in Figure 18.

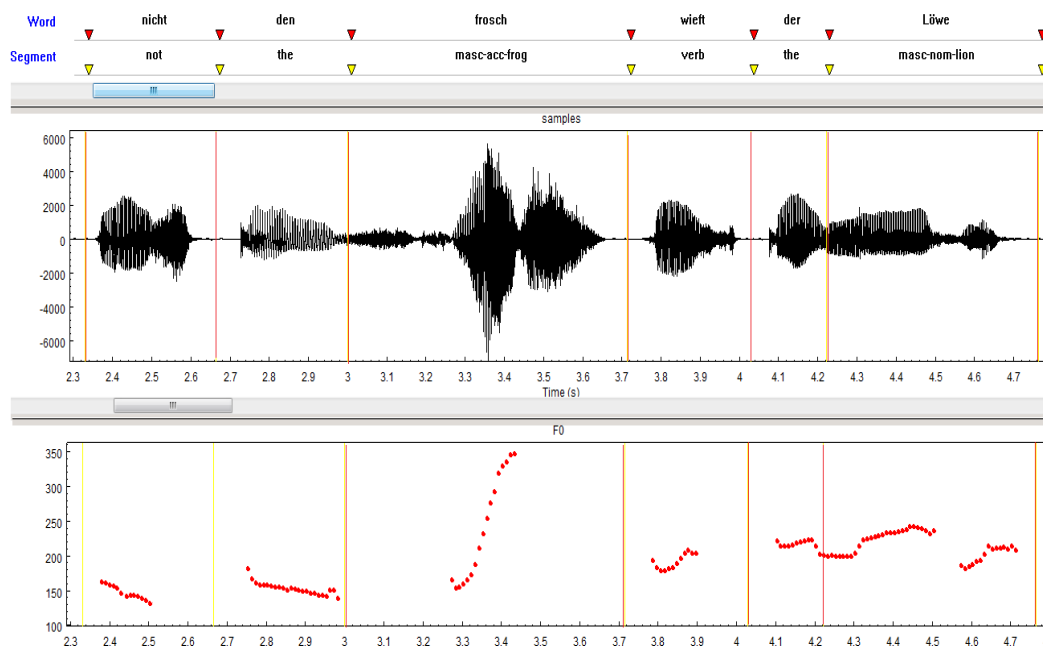


Figure 18: Example of the intonation of the first part of the correcting utterance as in sentence (20). The second part of the stimuli was recycled from the previous study (see Figure 13 and Figure 14).

²⁵ The second NP, printed in bold, was the same auditory stimuli used in the previous study. Except from that, all other auditory stimuli in this study were natural and were not manipulated.

The stimuli were recorded by the same female native speaker as in study 1. She was invited to utter the sentences as naturally as possible, leading to a L+H* accent on NP1. Other than the second part of the utterance (the target OVS-sentence), which was recycled from study 1, the speech material was not manipulated.

Procedure

The procedure of this study was the same as in Study 1 with the exception that the instructions did not come solely from one puppet but were embedded in a conversation between two puppets, as described above.

Pointing practice training & Word learning training: Pointing practice training & Word learning training were the same as in Study 1.

Film familiarization trials: Following the live enactment of the word learning training, the child then saw a familiarization trial for each verb in which he or she watched each of the two film scenes in sequence and heard the two puppets describing them. In this description, P2 was always wrong because he was too young to remember the novel verbs and was thus corrected by P1; for example:

P2: 'Guck mal, das heißt lemmen.' – 'Look, that's called lemming.'

P1: 'Nein P2, das heißt nicht lemmen, sondern wiefen. Das heißt wiefen.' -
'No, P2, that's not lemmen. That's weefing! That's called weefen.'

During these film familiarization trials, only one clip was visible on the screen while the other half of the screen remained blank. The side of the screen where the children saw the first picture (left or right) as well as the order of the novel words was counterbalanced across and within subjects. At the end of each film scene, the experimenter pointed to both animals and asked the child 'Wer ist das?' - 'Who's that?' The majority of the children had no problem spontaneously naming the participating animals. If a child did not name one of the animals, the experimenter told the child its name and asked him or her to repeat it, which nearly all of the children then did.

Test trial: The test trial procedure was the same as in study 1, except for the second puppet. At the moment where the attention-getter (the red dot) disappeared, P2 declared that he probably knows what happens in the next clips by saying a transitive SVO sentence, involving the novel verb and the right agent, but the wrong patient, as in (22). After finishing this sentence, the two clips appeared on the screen, accompanied by P1's prerecorded linguistic stimulus using the target verb in a transitive OVS argument structure, as in (23). After the videos had stopped, the prerecorded voice of the puppet asked the child to point to the correct (still) picture by asking, for example, "Zeig P2 das Bild: Den (+accusative) Löwen wieft der (+nominative) Hund!" – "Show P2 the picture: the (+accusative) lion is weefing the (+nominative) elephant". If the child failed to

point then the puppet repeated the question a second time, but she never asked the child to point again once she/he had already done so. Once the child had pointed to one of the two pictures, the next test trial began, preceded once more by the red dot.

Coding and Reliability

For every test trial, I coded responses for whether children pointed to the picture in which the post-verbal, second noun in the sentence was the agent. If a child did not choose either scene (3), I coded those trials as `wrong` (an alternative analysis in which these cases were excluded had no effect on the pattern of the results). For one participant, 6 trials were missing because of technical failure. In order to give all participants` data the same weight in the analyses, the remaining pointing values for this participant (=10) were coded as the total score (=100%) of this participant. All children were coded by the first experimenter, and an additional coder coded 25% of all trials for reliability, revealing a high agreement with the first rater (Cohen`s Kappa = 0.969).

7.2.3. Results and Discussion

I again tested for the proportion of times the NP occurring after the verb was identified as the agent of the action out of four. The data were analyzed using a 2 (Intonation) x 2 (Case Marking) repeated measures analysis of variance (ANOVA). There were main effects for both Intonation, $F(1,15) = 5.8$, $p = 0.029$ and Case Marking $F(1, 15) = 14.4$, $p = 0.002$, but no significant Intonation x Case Marking interaction ($F(1,15) = 1.13$, $p = 0.304$) (see Figure 19).

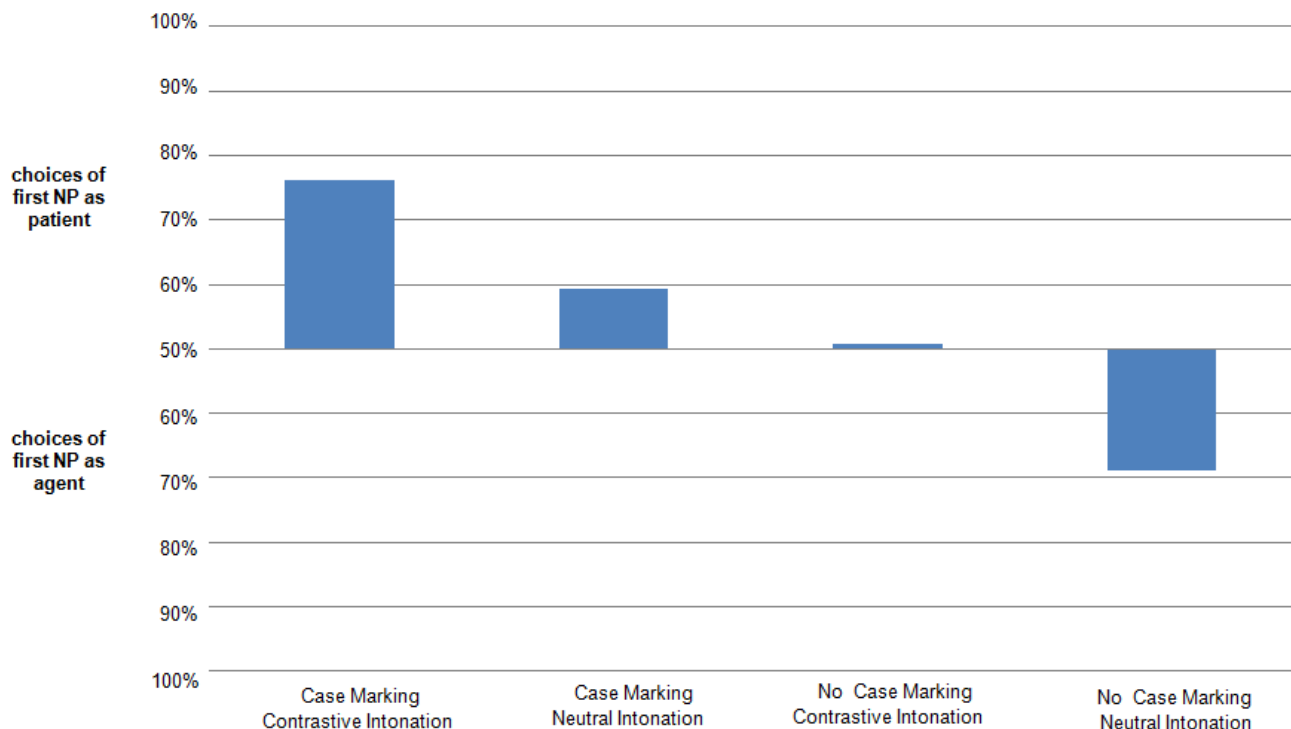


Figure 19: Results of the study in the four conditions. The diagram shows percentages of judging NP1 as either patient or agent as compared with chance, 50%.

Because the chance level for the dependent variable was always 50%, I also investigated in which conditions the children were above chance. The results show that the children were above chance in the Case Marking / Contrastive Intonation condition ($t(15) = 4.0, p < 0.001$) as well as in the Case Marking / Neutral Intonation condition ($t(15) = 2.2, p = 0.044$). In the No Case Marking / Contrastive Intonation, children chose agents and patients at chance level ($t(15) < 0.001, p = 1.0$), whereas children in the No Case Marking / Neutral intonation Condition relied solely on word order ($t(15) = -2.53, p = 0.023$).

A comparison between the two conditions Case Marking / Contrastive Intonation and Case Marking / Neutral Intonation revealed no significant difference (paired-sample t-test: $t(15) = 1.145, p = 0.270$), whereas choices in the two conditions No Case Marking / Contrastive Intonation and No Case Marking / Neutral Intonation revealed a higher judgment of NP1 as the patient, when this interpretation was supported by intonational stress ($t(15) = 3.0, p = 0.009$). These results strengthen and extend those of Study 1. In this study, children used natural intonation, as opposed to word order, in interpreting patient-first transitive sentences. In other words, children used a high pitched accentuation of the first noun phrase to identify a patient-first transitive construction. This effect was especially clear in the two conditions without case marking, which showed that intonation by itself, in the absence of case marking, is a sufficient cue for young

children to re-assess an agent-first interpretation. The two conditions with case marking, with and without intonation, did not differ, but they showed fairly high rates of success.

For a further analysis, we compared the results from the two studies presented in this Chapter (see Figure 20).

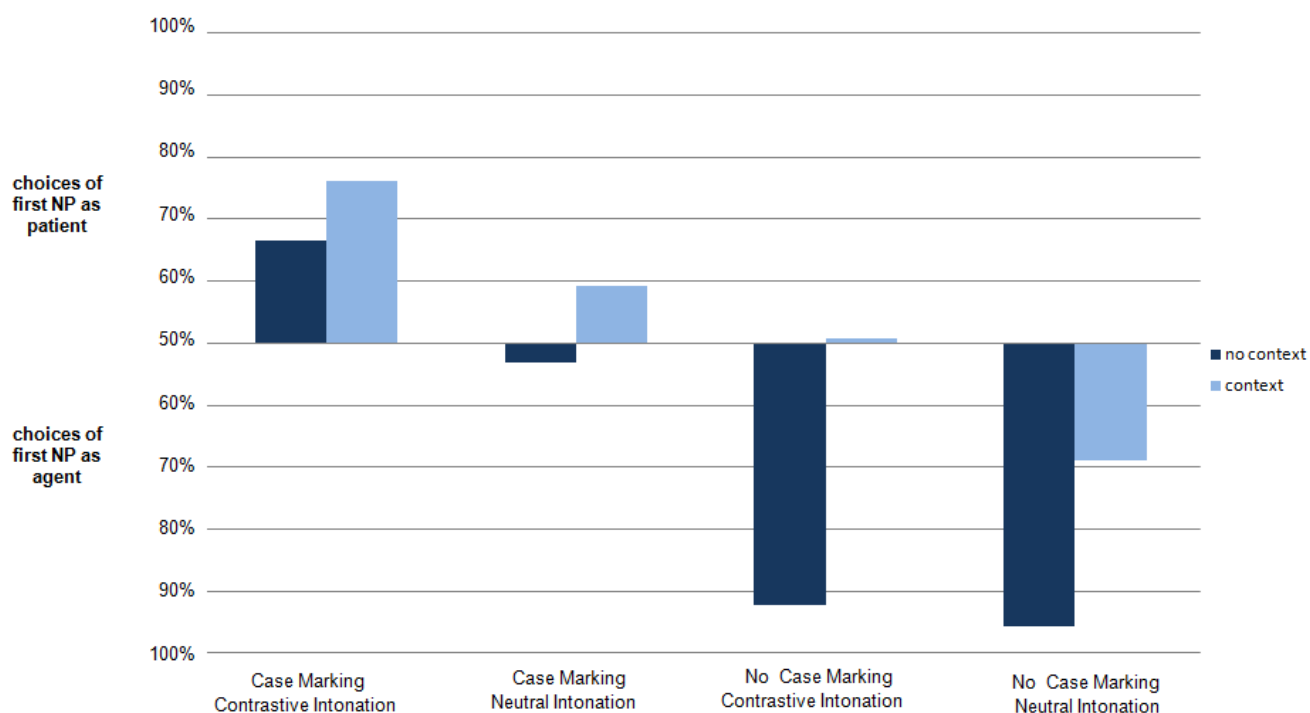


Figure 20: Comparison of results from the study presented in Chapter 7.1. (with no context) and from this Chapter (including context) in the four conditions. The diagram shows percentages of judging NP1 as either patient or agent as compared with chance, 50%.

The data were analyzed using a three-way mixed analysis of variance (ANOVA) with two repeated factors (Case Marking and Intonation) and one between-subjects factor (Context). There were main effects for both Intonation, ($F(1,30) = 10.7, p = 0.03$) and Case Marking, ($F(1,30) = 52.0, p < 0.001$), but no significant interaction between the two ($F(1,30) = 0.3, p = 0.541$). There was no significant interaction between Case Marking and Context ($F(1,30) = 2.5, p = 0.118$), or between Intonation and Context ($F(1,30) = 0.2, p = 0.602$), but I found a significant interaction between all three factors, ($F(1,30) = 4.4, p = 0.044$). A comparison between conditions of the two studies revealed no significant difference either in the conditions Case Marking / Contrastive Intonation (paired-sample t-test: $t(15) = 1.09, p = 0.285$), or in the two Case Marking / Neutral Intonation conditions ($t(15) = 1.72, p = 0.095$). Only those choices in the two conditions No Case Marking / Contrastive Intonation ($t(15) = 6.26, p < 0.001$) and

No Case Marking / Neutral Intonation ($t(15) = 3.16$, $p = 0.005$) revealed a significantly greater likelihood of judging NP1 as the patient, when this interpretation was supported by a combination of the prosodic pattern and the preceding context.

These results show the importance for children of a natural intonational realization in order to understand participant roles. Even in totally ambiguous constructions, the intonational form of an utterance can pull children away from their strong word order bias. The results from the study presented in Chapter 7.1. show that intonation is an important cue and helps children to understand agent and patient relations. But in isolation, without any help from other cues, the strong word order bias cannot be eliminated. If an appropriate context and intonational pattern are included (as for example that presented in this study), children can negotiate this bias and move towards ceiling levels when several cues are combined (i.e. case marking, intonation and discourse context).

7.3. General Discussion

In the current study I found that five-year-old German children recognize a high pitch accent on the initial noun phrase as a cue indicating a patient-first transitive construction. Thus, the prosodic cue is strong enough to pull children away from their strong word order bias whereby they interpret the first noun as an agent. In the study dealing with the role of intonation in resolving participant roles without context, as presented in Chapter 7.1., this effect could only be seen in combination with case marking. In those conditions where case marking was ambiguous, children, still fell back on their most reliable cue - word order. In the study where target sentences were presented in a more natural way with a combination of context and intonation, the results were strengthened because young children were using the intonational cue (in combination with case marking and context), as opposed to the competing cue of word order. In contrast to Dittmar et al.'s (2006) study, in which children of the same age systematically misinterpreted patient-first sentences, the children in these studies no longer depended on the most reliable cue - even in the absence of case marking. What this shows is that prosody has the power to work against this word order bias and that the information in the sound stream seems to be sufficiently rich to allow children to abstract participant roles.

The exact basis by which the children interpreted the prosodic cue remains as yet unknown. Focusing intonationally on certain words is a communicative function that serves to put emphasis on a particular part of an utterance. Varying widely across languages, it involves changes in duration, intensity, and vowel quality (e.g. Turk and White, 1999; Xu and Xu, 2005). However, the primary cue for perceiving focus is generally considered to be pitch

variation (Dahan and Bernard, 1996) and this was the cue that I controlled for in these studies²⁶.

Compared to Dittmar et al.'s (2006) results, the findings from the study presented in Chapter 7.1. are somewhat surprising. In the condition where case marking and word order contradicted each other, but none of the cues were reinforced by intonation (Case marking / Neutral Intonation), participants chose participant roles at chance, whereas the children in Dittmar et al.'s study relied primarily on word order. In my opinion, this is due to the natural mechanisms of speech, both psychological as well as physiological. In my study, intonation was computerized and manipulated and thus controlled; i.e. in the neutral intonation condition, children heard a completely flat intonation pattern, whereas Dittmar et al.'s children were tested with a task in which the experimenter uttered the target sentences in a live-situation. Even if the experimenter in that study had concentrated on a neutral vocal production, natural tendencies like declination or macro- and micro-prosodic cues provide a minimal prosodic pattern that the children could have used to decide on the agent and patient roles. In addition, the accusative marker in my study could have been more clearly articulated (due to intonational prominence) and thus more clearly perceived, as compared to Dittmar et al.'s study.

Dittmar et al.'s (2006) corpus study of input in six children recorded initially at 1;8 years and then again at 2;5 provides data for the frequency with which the types of sentences presented in my study occur around children in everyday speech. Overall, Dittmar et al. found 745 transitive sentences, 55% (410) of which had causative verbs. 21 % of those involved conflicting (but unambiguous) case marking and word order (my Condition 1). More interestingly, only 2 sentences in the corpus appeared with an object-first order and ambiguous case marking (my Condition 3). This means that although less than 1% of all causative sentences that children hear in the input are constructions containing non-canonical word order and ambiguous case marking, the prosodic characteristics of exactly the same constructions lead children away from a word-order interpretation in my study as presented in Chapter 7.2. In other words, despite the very low input proportions, children still manage to disambiguate these constructions when an intonational cue is present.

There are a number of possible explanations for these results, not necessarily mutually exclusive. It seems clear that the strong contextual cue provides the whole package in a more natural way and pulls the children towards an OVS interpretation. It is also possible that children could have learned the prosodic pattern associated with the patient-first transitive construction as a whole and abstracted a form-function mapping for the prosodic cue from the more frequent OVS causative constructions in the input which include case marking. However, the relatively weak results from the study without context (cf.

²⁶ For a discussion about the acoustic aspects of focus marking see for example Baumann, Mücke & Becker (2010) and Hermes et. al (2008).

Chapter 7.1.), especially in the conditions without case marking, would seem to argue against this hypothesis. It is also possible is that children are simply noticing an unusual prosodic pattern and are inferring that this suggests an unusual, marked interpretation, which they then need to guess from the various available options. One final possibility, which would provide even deeper insight into the acquisition of intonational meaning, is that children have come to understand more generally that new and —special information often stands in focus and receives prosodic highlighting. Thus it may be that by 5;0 children are in the process of abstracting a more general mapping from intonational prominence to sentential focus. This could be derived from simpler constructions. These might include utterances in which, while formally OVS, may well be learned as a whole together with their intonation (e.g. 'DAS mag ich' - 'that I like') as well as other syntactic constructions in which there is focal intonation such as imperatives ('Sitzt DA, nicht da!' - 'Sit THERE, not there').

In line with this view, Grassmann and Tomasello (2007) demonstrated in a recent word learning study that 2-year-olds already know that those words in an utterance that correspond to contextually new referents (and are thus "special" within the discourse) are prosodically highlighted (cf. Chapter 6.1.1.). And, this is also in line with the results from the studies presented in Chapter 6. This suggests that children interpret prosodic stress in language as being iconic of the speaker's intention to refer to a salient aspect of the situation. Interestingly, I have shown, as did Grassmann and Tomasello (2007), that only a combination of newness (or salience) and stress (or more precisely accent) together were effective. In the study with context, where children used the prosodic cue much more effectively, the first noun phrase referred to the new participant in the situation, and critically, the contrast was with a participant who was the patient in the preceding discourse context. Furthermore, the linguistic material that is new, or in some sense contrastive, was prosodically highlighted compared to given or contextually available information. Indeed, it is not totally clear that these are separate hypotheses, as it is possible that even adult Germans use the intonation typically associated with patient-first transitive sentences in this more general way, rather than as part of the transitive construction as a whole.

In order to resolve syntactic ambiguities, children need sentences that contain multiple cues - according to Bates and MacWhinney's (1987) concept of coalitions-as-prototypes. What this means is that because sometimes several cues may indicate the same function—providing extra information—children should find it especially easy to comprehend prototypical transitive sentences, e.g. with both word order and case marking (and perhaps other cues) working in coalition. This study adds the fact that children do not just use morphosyntactic cues like word order and case marking to disambiguate participant roles. Prosody, especially in combination with an appropriate context, is an important cue which in the absence of clear morphological cues can modulate subject and object assignment. Thus, the problem of processing sentences with non-canonical word order can be partially alleviated when these utterances are presented with the appropriate intonation and the appropriate context. In their early development children can only interpret sentences which contain

combinations of cues in the most frequently heard patterns. However, development consists in starting to identify the separate contribution of each cue. The present study indicates that, in line with Usage-Based approaches, both the context and sentential intonation should be treated as cues of considerable importance and investigated as such. It is likely that intonation interacts in complex ways with a number of different morphosyntactic cues, and indeed I provide some evidence for this possibility. In some cases the prosodic pattern may be a part of the construction itself, whereas in other cases it may be being used more generally, for example as a contrast, in order to stress a particular noun phrase which then triggers a specific interpretation of a particular construction. But again, this may be a false dichotomy, as in many cases the distinction between these two interpretations is unclear - a good example being the English cleft construction, for example, "It was the DOG that got sick"; in this case the stress on *dog* could be interpreted by either route. In any case, the larger point is that to fully understand young children's skills at interpreting sentences online, the role of intonation and context must be taken into account.

Part III: Empirical Studies - Production

8. Young children's intonational marking of new and given referents

8.1. Introduction

According to the Usage-Based approach to language acquisition, it is of particular importance in the language acquisition process to understand another's intentions (cf. Chapter 3.2.). With this understanding, one can interpret the communicative intentions of other persons. One instrument that offers the possibility to convey communicative intentions is intonation. Thus, for young language learning children, it is not only of particular importance to understand intentions by intonation, but also to produce the appropriate intonational pattern in order to make their own intentions understandable. By accenting certain words or phrases within an intonation, a speaker signals a certain state of newness (and importance) for that particular word or phrase. In contrast, the lack of accent (deaccentuation) is said to signal Givenness to that part of the utterance. Intonation is therefore an important instrument in order to organize the informational status of target referents within an utterance and to optimize the conveyance of information. Thus, intonation is related to what a speaker knows or thinks she knows is present in the hearer's mental world. And, entities in this mental representation can be manipulated with regard to the hearers background.

Overall, it is typically assumed that in West-Germanic languages like English or German the placement of pitch accent is crucial for the marking of the informational status of referents. For Halliday (1976), the distinction between given and new information is central to the pragmatic analysis of utterances. He interprets new information as "the main burden of the message" (1967b: 204), marked by the nuclear pitch accent²⁷. The nuclear pitch accent is placed on exactly that part of the utterance to which the speaker attributes the function of 'new' to. In Halliday's understanding of the concepts of given and new information therefore, the choice of a particular pitch accent seems to be a very pragmatically one because the speaker chooses a certain intonational realization for a referent, based on his intentions. For example, accenting a referent often indicates that new information is introduced into the discourse, whereas deaccenting may be used in the case of already established (given) information (e.g. Ladd, 1996, Gussenhoven, 1984). Additionally, accentuation is typically used to signal a contrasting relation between referents.

Several scholars find this classification between accented vs. unaccented for new vs. given information insufficient and have gone beyond such a binary distinction. As already mentioned in Chapter 2.3.1., Pierrehumbert & Hirschberg

²⁷ Halliday uses the term tonic component of the tone group, which corresponds to the (nuclear) pitch accent in AM-theory

(1990) propose for English that not only deaccentuation, but also different pitch contours containing a low Pitch Accent (L*) indicate that the speaker does not want to add something new to the mutual beliefs of the speaker and the hearer. Thus, L* accents – in addition to deaccentuation – seem suitable to mark given information. Contours containing a high pitch accent (H*) are assumed to signal newness. According to this, Baumann & Hadelich (see Baumann, 2006) in a perception study manipulated the intonational realization of utterances concerning their informational status and asked German adults to judge the appropriateness of the used accent types. The results showed that H* was perceived to be the most appropriate marker for new referents. For given referents, listeners judged deaccentuation as most appropriate, whereas H* was least acceptable. These results indicate that German native listeners are sensitive to the degree to which a referent is given within a discourse, and that they have intuitions about the intonational marking, which go beyond the dichotomy of accented vs. deaccented. Thus, the speaker is in fact sensitive to what cognitive status a referent has in the mind of the listener – and vice versa. And, both participants of a conversation understand what a particular intonational pattern means, i.e. new information requires a certain effort whereas given information does not. This is important because in order to understand the intention of a speaker, the hearer has to know how to read that particular realization.

In terms of infants and young children's understanding of intentions conveyed by intonation, several studies have shown that they understand what others do and do not know and about what is given and new to people in a particular situation (cf. Chapter 4.1.). Additionally, as we have seen in the comprehension studies presented in this thesis, children do understand that certain intonational patterns are important for understanding what others intend to say.

However, it is yet unclear whether young children, who have only recently entered the multi-word stage, can use this knowledge about what is new and given for another person in their own intonational realization. In order to understand the process of the acquisition of language, the answer to this question is of particular importance. The use of the appropriate intonational pattern is an important developmental step and it is of essential importance to convey its own communicative intentions in order to be understood. Whereas the intonational encoding of the cognitive status of target referents in adults is widely examined, evidence about children's competence in this area is scarce. However, intonation, as referring to the patterning of pitch changes in utterances, is commonly assumed to be an early-developing component of language and to be mastered by children more or less before they produce their first words (e.g., Lewis 1951, Bever et al., 1971, Crystal 1979, Locke 1983). This belief is consistent with theories positing that intonation is physiologically or emotionally "natural".

Overall, in terms of young children's use of intonation in order to mark the information status of target referents, it is typically assumed that children accent new, but not given information in their own speech (e.g. Wieman, 1975;

MacWhinney & Bates, 1978; Baltaxe, 1994). However, as already stated in Chapter 4.2., most of the studies that examine the use of intonation in order to mark the informational status of discourse referents have not looked at spontaneous data or tested children that were more experienced with language. Moreover, none of the cited works provide any detailed or useful phonological or phonetic analyses. Instead, stress is used as an equal term for all kinds of accentuation. As a result, nothing is known about the relationship between types of pitch accent (including deaccentuation) and the according cognitive representation of that referent, or other prosodic features in young German children who have just begun multi-word usage. In order to fill this gap, I systematically investigated young German children's intonational marking of the informational status of discourse referents in semi-spontaneous speech. Here, the intonational realization of given target referents is of especially importance. In order to realize the intonational form of such a target referent, it is necessary to understand its cognitive representation not only in its own mind, but particularly in the mind of another participant in the communicative act.

8.2. Data & Method

Using a story-telling task, 2;6 and 3;0 year old children were asked to describe four different picture books in which the occurrence of a target referent was manipulated: it was either inactive (and thus new) or already established into the discourse (and thus given). Additionally, in one case, the target referent was manipulated in such a way that the child had to utter a correction in a contrastive way. The question was whether children have already established the ability to mark the difference between new, given and contrastive target referents by intonation. The second question I sought to answer was in which way the new and the contrastive element prosodically differ from each other. To answer these questions, I analyzed the use of different types of pitch accent with which the informational status of target referents were realized. Furthermore, differences in the prosodic realizations of these elements, namely pitch range, was investigated. Additionally, the data was compared with that of adults which were tested in the same method.

Participants

Sixteen 2;6-year-old children (range 2;6 – 3;0, mean = 2;7; 6 boys and 10 girls), sixteen 3;0-year-old children (range 3;0 – 3;6, mean = 3;3; 8 boys and 8 girls) and eight adults were included in the study. All participants were monolingual German and were born and raised in the same dialectal environment. For the 2;6 year-olds, one additional child was tested, but excluded from the study because less than 50% of the target referents were uttered; for the 3;0 year old age group, four additional children were tested but excluded from the study because they either showed disinterest in the picture books (1) or uttered only 50% or less of the target referents (3). Children were recruited from a

database of parents from diverse socio-economic backgrounds who had volunteered to participate in psychological studies. All children were tested in nursery schools in a medium-sized German city; all adults were tested in a sound-proof room. In order to test the ability to comprehend and to produce sentences, an additional 50 % of the 3-year old-children took part in a language development test (SETK 3-5; Grimm, 2001). Two subtests were conducted. In the subtest "Verstehen von Sätzen", the children received a comprehension task, in which they should solve different task with different objects (e.g. "Put all red buttons in the box"). Here, the children who participated in the test had a mean range of 56 (rang 46 -64). Additionally, children received the subtest "Enkodierung semantischer Relationen", in which pictures should be described. In this task, the children who participated had a mean score of 55 (range 41 – 79). The mean scores were, therefore, as the expected ones for their age range (expected: 50, SD 40–60).

Materials

Four picture books were designed, all with a similar concept in which a target referent was presented in one of three informational contexts: (1) new, defined as information conveyed by a referent that was not previously mentioned or indirectly touched upon (e.g., via semantic relatedness), (2) given, defined as information conveyed by a referent that was mentioned previously in the discourse, and (3) contrastive, defined as a correction or protest to a preceding incorrect referent.

Four target referents were chosen. These were: 'Möwe' – 'seagull', 'Biene' – 'bee', 'Eule' – 'owl', and 'Igel' – 'hedgehog'. These target referents were chosen in order to fulfill certain criteria: in order to get as much speech material as possible; they should be child-friendly and be well known by young children²⁸. In addition, the target referents should be disyllabic with a sonorant segmental make-up to facilitate pitch analysis. And, the referents should not switch form when declined.

All four picture books contained 6 pictures. Picture 1 was intended to introduce the topic (e.g. a forest). Picture 2 introduced the target referent (e.g. a hedgehog). Picture 3 introduced a distractor referent (e.g. a deer) with the target referent visible in the background of the picture (in order to keep the target referent active). In picture 4 and 5, the distractor referent acted on the target referent in a causative way (e.g. the deer is washing the hedgehog). The action was chosen in order to elicit a transitive SVO sentence in which the target referent was mentioned as the patient. On the last picture, the target referent left the scene. Thus, picture 2 tried to elicit a verbal production of the target referent in a 'new' form, picture 3-5 in 'given' form and picture 6 attempted to elicit a

²⁸ According to the German CDI (Szagun 2009), all target referents except from *Möwe - seagull*, were known by 2;6 year old German children.

‘contrastive’ utterance of the target referent as a correction of the experimenter’s incorrect naming. Appendix C shows an example of one of the picture books.

Design and Procedure

I tested all children and adults with four different picture-books using a story-telling task. During the session, the child and the experimenter sat in a comfortable position in a quiet room at their nurseries. The adults were tested in a soundproof room at a table. In the test trials, participants were presented with one picture book after another involving one of the four target referents. The participants were asked to describe the picture-books. During the test-phase for the children, the experimenter said as little as possible but made sure that the discourse did not stop; for example, by helping to keep the plot moving. All participants received each of the four picture-books in one session. The order of the picture books was counterbalanced in a 4*4 Latin square.

The test session lasted for approximately 20 minutes. All sessions were audio-recorded with a digital microphone (Olympus LS-10) which was positioned approximately 50 cm in front of the child. Additionally, all sessions were videotaped with a camera in front of the child.

Warm-up: The aim of the warm-up phase was to familiarize the child with the situation and the task: namely, to talk about different objects and pictures. To do so, the experimenter introduced a ‘surprise-bag’ with 8 different items (e.g. a toy dog, a toy helicopter). The child and the experimenter took turns taking items out of the bag and talking about them. If necessary, the experimenter encouraged the child to talk more about the item by asking several questions, for example, “Do you have a dog?” “What’s his name?” “Do you go out with him very often?...” The experimenter made sure that the child engaged as much as possible in this conversation.

Practice phase: After the warm-up phase, the experimenter told the child that he wanted to show some pictures he had made. These pictures contained different single items (= 7 pictures), including pictures of the target referents and distractor pictures (e.g. a duck). Pictures of target referents were different to those used in the test trials. By showing these pictures, the experimenter could test whether the children knew the words for the target referents and, if necessary, correct or teach the words. Additionally, the experimenter showed 10 pictures on which animals (different from the target referents) enacted transitive actions on each other. By doing this, he could make the child familiar with uttering full transitive sentences. For each of the pictures, the experimenter asked the child to describe the picture and, when necessary, he helped out.

Test phase: After the practice-phase, the experimenter wanted to show a “real” picture book to the child. The children were asked to describe the story in the picture-books. While watching the books, the experimenter said as little as possible in order to let the child tell the story. When necessary, the experimenter encouraged the child to talk by describing the background scene (e.g. ocean, meadow), but he never used the target referents in the discourse (instead, he

only talked 'around' the target referent, e.g. the wings of the seagull, the coat of the owl...) ²⁹. If the child used a pronoun rather than a full NP to describe the target referent, the experimenter named the target referent in order to activate it. In order to elicit a contrastive utterance from the child, the experimenter described the last picture of each book by saying: "Look, X is running away!" Here, he used an incorrect referent, for example 'cow'. Each child was presented with all four picture stories in a counterbalanced order.

The test phase for the adults differed slightly to the children's test procedure because they did not get a warm up and practice phase. Instead, adults started directly with the test-phase in which they were asked to describe the picture books to the experimenter. Participants received no information about what quantity or quality the picture-book descriptions should have. Instead, participants were asked to speak at their own speed.

Coding and Reliability

For every picture-book description I separated those intonational units in which the target referent occurred (for examples of the utterances that participants from each groups gave in each of the three conditions, see Appendix D). Only natural and spontaneous realizations of a target referent were analyzed, i.e. not answers to a question or in cases in which the target referent was uttered as a pronoun. The target referent that the participant uttered first within the discourse was coded as 'new'. The referent that was uttered after this activation of the target referent (either by a spontaneous realization or by activation of the experimenter) was analyzed as "given". For the contrastive analysis that realization of the target referent that was uttered as a protest after E's wrong labeling was analyzed as "contrastive".

Due to problematic with eliciting spontaneous speech from young children ³⁰, the primarily question at this stage of the study was whether or not the participants would utter the target referent in the three conditions. Thus, I checked whether and in how many cases the participants realized the target word within the three conditions (see Table 9).

²⁹ It is important to note that the experimenter took care about an ongoing plot of the stories within the picture books. In this sense, the task was not just an object-naming task but rather a story-telling task.

³⁰ Problems that can arise with young children are for example their shyness e.g., they do not want to talk to strangers, they do not know the target referents or children are unaccustomed to the procedure.

Table 9: Number of possible realizations of the target referent (for children = 64, for adults = 28) for the three age groups and their actual realizations (absolute and relative).

	2;6		3;0		adults	
New	57 / 64	89,06 %	53 / 64	82,81 %	26 / 28	92,87 %
Given	56 / 64	87,5 %	51 / 64	97,68 %	27 / 28	96,42 %
Contrastive	59 / 64	92,84 %	52 / 64	92,81 %	27 / 28	96,42 %

This table shows that in all age groups, target referents were produced in at least 80 % of all cases. This made it possible to make a reliable analysis of the intonational realization of target referents within a discourse throughout age-groups and conditions.

In order to make sure that the participants did not treat the task as an object-naming-task, in which the target referent was uttered by using a bare noun phrase (NP), e.g. "A seagull!", but rather as a story-telling-task, I analyzed the syntactic structure of the utterances in the three conditions.

This is especially important because the intonation possibilities are quite different for NPs vs. sentences. In particular, deaccenting is impossible by definition in simple object naming. However, because it is not possible to realize an Intonation Phrase with no pitch accents at all (this is basically definitional). If younger children, due to poor speech performance are more likely to produce IP's containing only one accentable referent (like in "A seagull!"), then it falls out automatically that a lower percentage of their productions involve deaccented referents. Table shows percentages of cases in which the target referent was uttered by the use of either a NP (e.g. "A seagull!") or by the use of a whole sentence (e.g. "The boy is feeding the seagull!"). Figure 21 shows percentage of cases in which participants from each group used a NP to utter the target referent or a whole sentence.

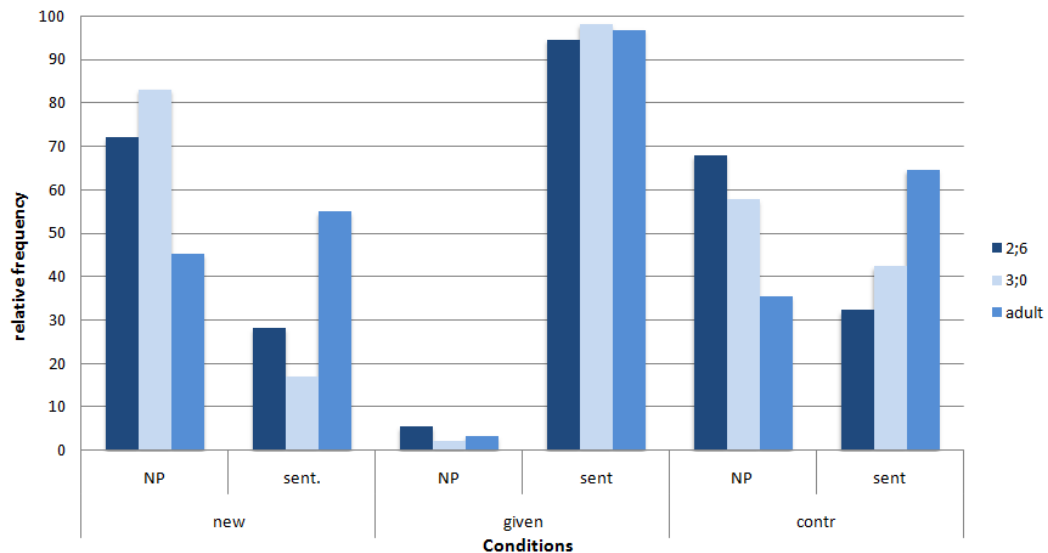


Figure 21: relative frequency in percentage of cases in which participants from each group used either a NP (e.g. "A seagull!") for the realization of the target referent or a whole sentence (e.g. " The boy is feeding the seagull!").

As Figure 21 shows, adults used in 55% of all cases a sentence to carry out the target referent in new form, whereas both child-groups did so in less than 30 % of all cases³¹. However, the focus from this study lies in the intonational realization of referents that already are established within a discourse. As we can see from the previous Figure, all age groups realized the target referent in this condition in more than 95 % off all cases by uttering a whole sentence. Thus, a reliable analysis of the intonational realization in this condition can be done.

In order to carry out the prosodic annotation, the recordings were digitized and annotated using the EMU Speech Database System (see Cassidy & Harrington, 2001; and <http://www.sourceforge.net/>). EMU is a collection of software tools for the creation, manipulation and analysis of speech databases. It can display various tracks such as the speech waveform, a spectrogram, the F0 contour and several layers for different kinds of labels, which can be arranged in a sequential or hierarchical order. The annotation followed the conventions of German – ToBI (cf. Chapter 2.3.1.). Using this framework, the intonation unit containing the target referent was segmented at the level of the syllable using information from a wide-band spectrogram. Additionally, the onset and offset of the lexically stressed syllable was marked. Following this, position and value of local F0 maxima (max) and minima (min) were measured in order to describe the intonational pattern, that is high pitch accent (H*), low pitch accent (L*) and

³¹ It has to be noted that the strategy to utter the referent in a bare NP is absolutely sufficient as this is the only new referent in the picture. As Grice (1975) pointed out in his Maxim of Quantity: "Make your contribution as informative as required." and "Do not make your contribution more informative than is required" (1975:1) (see also Salomo et al, 2010).

deaccentuation³². The domain in which these landmarks were set consisted of the lexically stressed syllable, the preceding syllable and the syllable following it. With the same measurements it was also possible to analyze the pitch range with which the target referents were realized³³. Figure 22 shows an example of the F0-contour of an example in the given condition with the regarding landmarks of the F0-minimum and maximum and the landmarks.

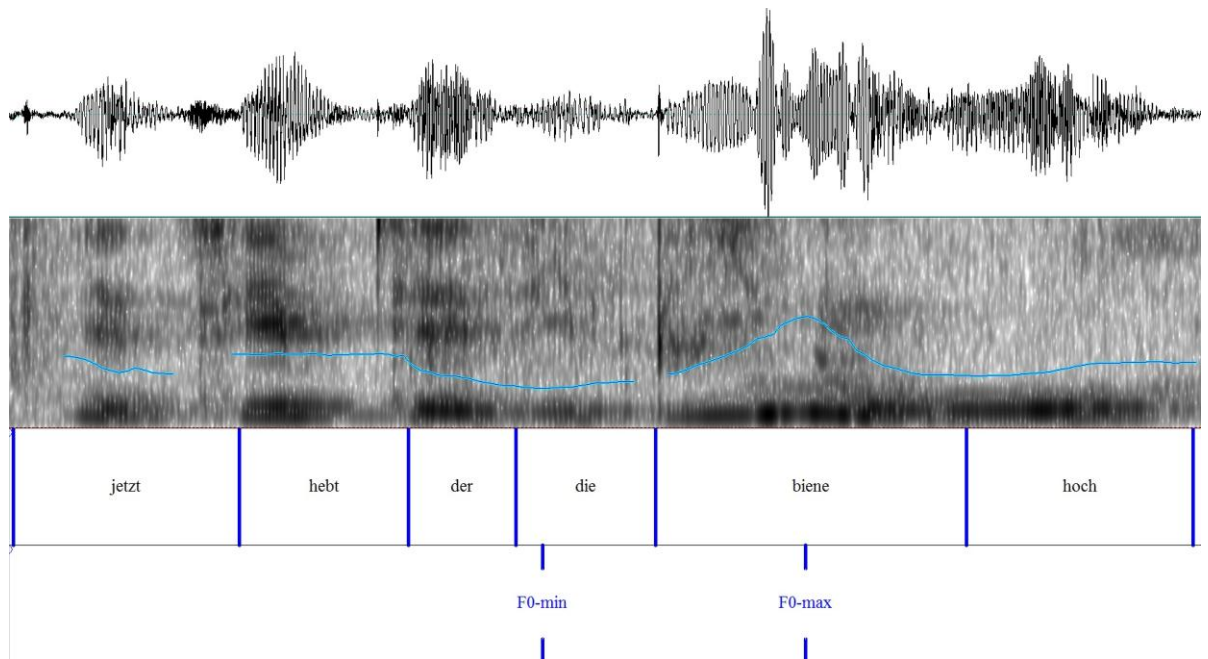


Figure 22: Example display of the realization of the target word "biene" in the given condition. The first row of the example shows the oszillogram, the second row the spectrogram and the fundamental frequency of the utterance "jetzt hebt der die biene hoch" – "now he takes up the bee". The third row shows the position of word boundaries, the fourth row the position of the local F0 maxima and minima. To do so, the lexically stressed syllable, the preceding syllable and the syllable following it were taken into account.

³² Please note that all possible intonational contours in German (see **Table 4**) are subsumed under these categories. This means that all intonational contours containing a high pitched accent (e.g. L+H*) were categorized as H* and all intonational contours containing a low pitch accent (e.g. L*+H) were categorized as L*.

³³ In order to analyze differences in the prosodic realization of the target referents, several additional measurements are possible. For example, the length of a target referent gives sufficient information about the effort that is used to realize it. But, the length of words depends on their position within an utterance. Due to physical characteristics of the speech signal, utterance-final words tend to be longer (known as final-lengthening) (see Beckman & Edwards, 1990). However, because this study examined the prosodic realization of target referents within spontaneous speech, the occurrence of a target referent could only be semi-controlled.

All realizations of the target referent were coded by a ToBI expert. An additional phonetically naïve listener was trained in EMU and GTobl. After this training, he coded 25% of all trials for testing reliability (= complete session of four randomly selected children and two randomly selected adults). The second coder had no information about the context of the utterances, the condition to which the target referent belongs or the judgments of the first judge. This reliability judgment revealed a high agreement with the first coder (Cohen's Kappa = .831). For cases of disagreements, the first and the second coder analyzed and discussed them together, leading to a perfect agreement in all cases.

Statistical Model for Main Analysis

Since the response variable was binomial (participants responded with one of three accent types yes/no) and since there were repeated observations of the same subjects, I used a Generalized Linear Mixed Model (GLMM) (Baayen, 2007). Into this I included as fixed effects the covariates condition and group, and as random effects subject and word. In principle such an analysis is somewhat similar to repeated measures ANOVA. However, it also permits to analyze a binary (i.e. yes/no) response variable. In addition, it can account for more complex structures of random effects, i.e. allowing for more than a single blocking factor (like 'individual' in a repeated measures ANOVA) and also crossed blocking factors (i.e. target referents and individuals, with each individual tested with each target referent). I fitted the models in R (version 2.8.0; R-Development-Core-Team, 2008) using the function `lmer` of the package `lme4` (Bates, Maechler, & Dai, 2008), with binomial family, logit link function, and maximum likelihood fitting. I tested for significance using likelihood ratio tests (Dobson, 2002) whereby I compared the fit of a full model with that of a corresponding reduced model using the R function `anova` with argument `test = "chisq"`. I first established the significance of the global model by comparing the fit of the full with that of the null model comprising only the random effects. I then tested the significance of the interactions, beginning with the three-way interaction and removed interactions when they were not significant (but only when they were not included in a higher order interaction which was kept in the model because it was significant).

8.3. Results and Discussion

Pitch accent type

I looked at the mean proportion of times children used one of the three accent types H*, L*, or deaccentuation, in each condition (see Figure 23).

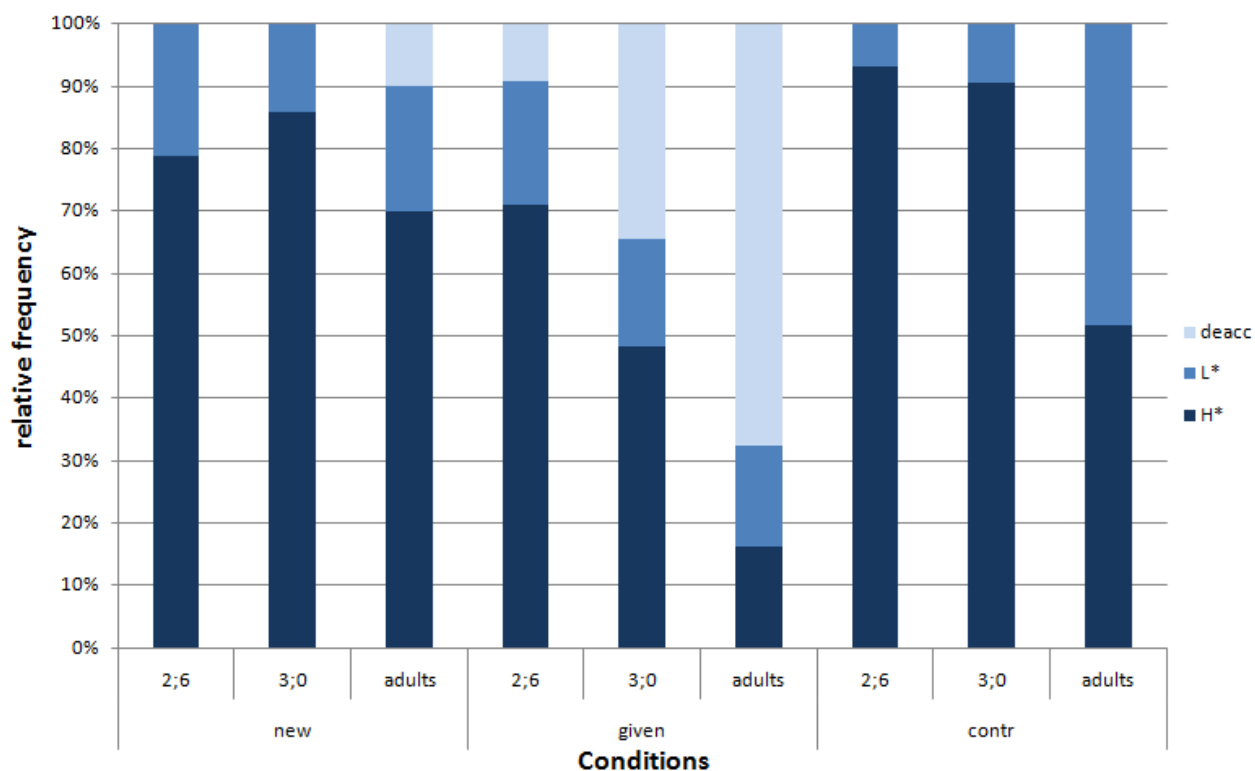


Figure 23: Results of the Pitch Accents with which the target referent was realized in the three conditions. The diagram shows percentages of the use of one of the three accentuation types.

In a first test, I analyzed the use of the H* pitch accent. Statistical analysis of the data revealed that the full model (including condition, group, and their interaction and the random effects) was clearly better than the null model (including only the random effects; likelihood ratio test: $\chi^2 = 65,95$, $df=8$, $p<0.001$). Furthermore, I found a marginally non-significant interaction between group and condition ($\chi^2=8,6$, $df=4$, $p=0.07$). This suggests that the use of the high pitch accent is mainly manifested in the responses from the adults, but not in the children groups. Thus, there is no significant interaction but a tendency between the use of the pitch accent H* and condition and group. Post-hoc tests that were conducted as mixed models support this hypothesis. Within-group analyses about the use of one of the three pitch accent types revealed a significant difference for the use of H* in adults ($z=3.98$, $p<0.001$) as well as for

the older children ($z=3.58$, $p<0.001$) but not for the younger children ($z=1.133$, $p=0.25$).

Comparing the use of H* between groups, there was a significant difference of the use of H* in adults choices in the given condition compared to both other groups (vs. older children: $z=2.148$, $p=0.032$; vs. younger children: $z=3.944$, $p<0.001$). Additionally, the younger children realized the target referent in the given condition significantly more often with a H* than the older children ($z=2.078$, $p=0.03$).

The same analysis was made for the use of the low pitch accent, L*. Again, statistical analysis of the data revealed that the full model (including condition, group, and their interaction and the random effects) was clearly better than the null model (including only the random effects; likelihood ratio test: $\chi^2=21,69$, $df=8$, $p<0.005$). Further I found a significant interaction between the low pitch accent, group and condition ($\chi^2=12,3$, $df=4$, $p=0.01$). Post-Hoc tests revealed no significant values in any of the between or within group effects.

For the use of deaccenting target referents, statistical analysis of the data revealed that the full model (including condition, group, and their interaction and the random effects) was clearly better than the null model (including only the random effects; likelihood ratio test: $\chi^2=129,93$, $df=8$, $p<0.001$). Further I found no significant difference between the use of deaccentuation and condition and group ($\chi^2=1,29$, $df=4$, $p=0.86$). Post-Hoc tests for deaccentuating the target referents revealed that only adults differed significantly between the two conditions 'new' and 'given' ($z=-4.25$, $p<0.001$). Additionally, adults choices of using this kind of realization differed significantly to those of the other two age groups (vs. older children: $z=-3.549$, $p<0.001$; vs. younger children: $z=-2.694$, $p=0.007$). And, the older children deaccented the already established referent significantly more often than the younger ones ($z=2.694$, $p=0.007$).

Pitch range

To analyze the use of pitch range, I measured the local min and max of the fundamental frequency in Hertz. Because the Hertz scale is linear whereas the perception and production of pitch is not, it was necessary to calculate the difference between the min. F0 and max. F0 in semitones, using a logarithm ($39,863 \cdot \text{LOG}(\text{max}/\text{min})$). The data was log-transformed (family = gaussian, link = identity) and tested with a likelihood-ratio test. Analysis of the data revealed, overall, that condition and group significantly explained the differences in pitch range ($\chi^2=65.067$, $df=8$, $p<0.01$). Following on from this, I did a check of assumptions by a visual inspection of residuals plotted against predicted values. The data was then analyzed using a generalized linear mixed model (random effects = subject and word; fixed effects = condition * age). P-values were obtained using Markov-Chain-Monte-Carlo sampling (MCMC). These analyses revealed a significant effect of age (mcmc; $p<0.001$) as well as for condition (mcmc; $p<0.001$). Comparing the data concerning the within-group differences between the conditions 'given' and 'new' revealed a significant difference between the pitch range adults used to mark the target referent in 'new' and

‘given’ form (mcmc; $p < 0.001$) as well as for the younger children (mcmc; $p = 0.014$). This is in fact not surprising for the adults, as this group realized the target referent in 86 % of all cases by deaccentuation (resulting in a narrower pitch range). Interestingly, the pitch range for the younger child did differ significantly, although this group realized both new and given target referents with a similar amount of high pitch accents. However, for the older children, no significant difference could be found (mcmc; $p = 0.262$) (see Figure 24).

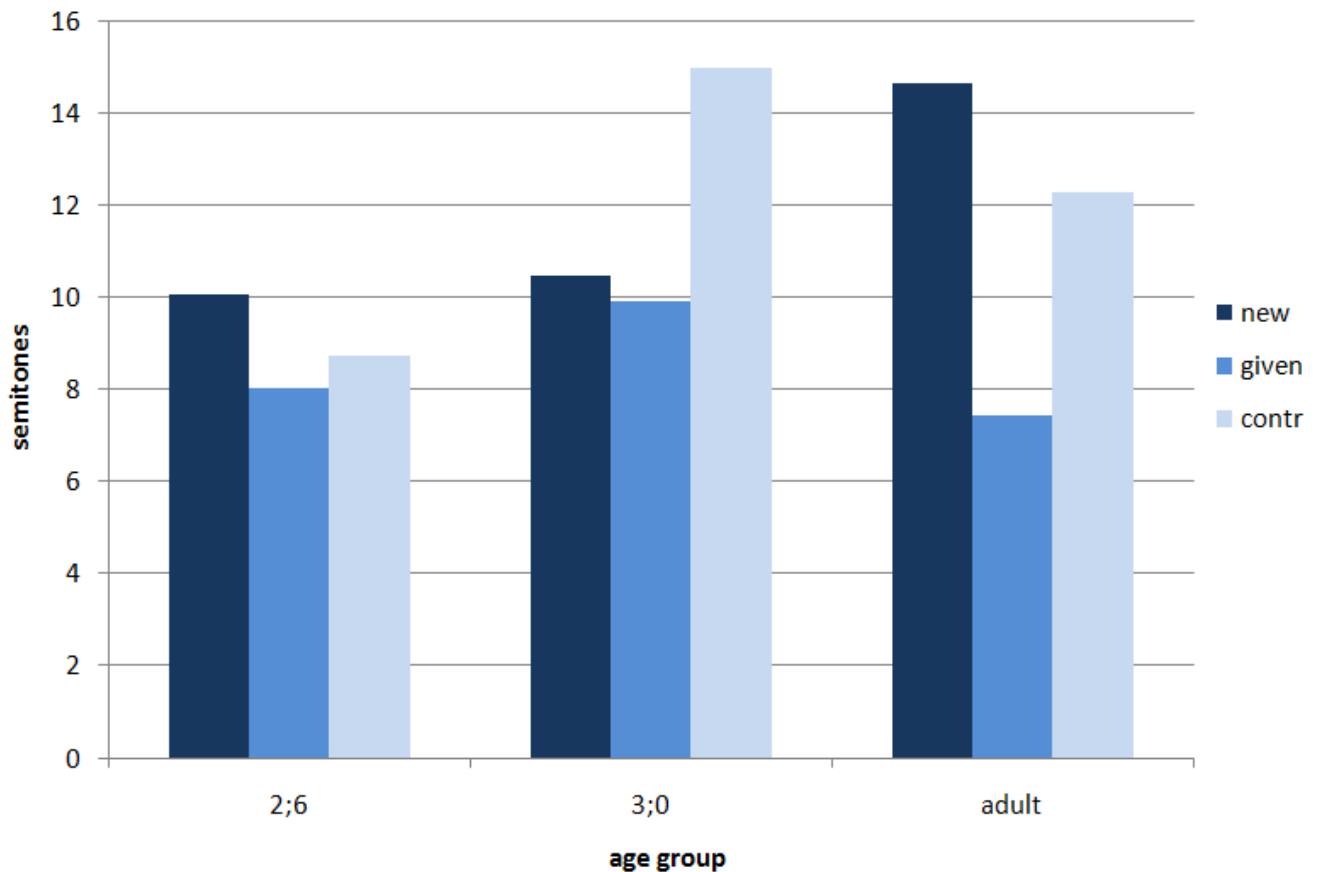


Figure 24: Results of the pitch range with which the target referent was realized in the three conditions. The diagram shows the realizations of the target referents in semitones.

What these results show is that adults as well as children in both age-groups behave similarly in realizing information that is newly introduced into a discourse. That is, young children already understand that information that is not recoverable from the preceding discourse or that is newly introduced need to be highlighted. Equally, to correct a proceeding referent that is incorrect, both child groups mainly use a high pitched accent for contrast, whereas, with respect to the energy used to do this, the older children put much more effort into the correction. However, whereas I could confirm Baumann’s (2006) results that adults tend to de-accent given information, I found that the younger children do

not. Instead, they treat given information as if it was new by accenting it. This is consistent with findings from Chen (in press) who found that 3-year olds produced more deaccented tokens than 2-year-olds.

The question is thus why children, who are just entering the multi-word stage, do not deaccent given information. There are three obvious hypotheses: First, younger children do not understand that the second target referent mentioned is old information. However, this explanation seems unlikely. As we have seen in the previous chapters, infants at the age of 14 month already know what is new and given for another person. Second, younger children do not have sufficient control over their speech-organs at this stage. This hypothesis is supported by the fact that young children put the same energy into the realization of new, given, and contrastive information, whereas older children put more effort into correcting someone's incorrect naming. Thus, children seem to "learn" more about the usage of their speech organs. The third - and not mutually exclusive hypothesis - is that children could have learned their intonational behavior from the input. Accenting given information is a characteristic of the motherese speech register used by most western, middle class parents. From an acoustic point of view, motherese has a clear signature (high pitch, exaggerated intonation contours) and has been shown to be preferred by infants over adult-directed speech and might assist infants during the language acquisition process (Kuhl 2004). Thus, the nature of the speech directed to children could play a major role in their learning of the conventional forms of intonation realization to express informational status.

However, all three hypotheses involve a certain developmental aspect and are supported by the findings that older children behave in a more adult-like manner. Thus, the usage of appropriate intonational behavior seems to develop with age. But, it seems that there is no easy answer to the question of exactly how children learn how to use intonation in an appropriate way. What we know from previous studies is that children at 9 months of age do know what others know. But, as the results suggest, it seems that children have difficulties translating this knowledge into intonation. This could be due to articulatory difficulties which seem to disappear by preschool age, as found by deRuiter (2010). However, in order to find out more about the influence of the input, i.e. the speech young children are supposed to use in everyday life, further research is necessary. The question of what influence of the input and its effect on young children's intonational development will be dealt with in the following chapter.

9. The role of the input for children's intonational development

9.1. Introduction

When talking to their children, adults use a different kind of language as compared to adult-adult speech. These differences are mainly characterized by the use of shorter sentences, including longer pauses as well as a change in the prosodic characteristics of their speech (e.g. Fernald & Simon, 1984; Fisher & Tokura, 1995). Additionally, speech to young children has higher fundamental frequency, greater F0-variability and expanded F0-range including more prosodic repetition (e.g. Fernald & Simon, 1984; Papousek, Papousek & Haekel, 1987; Fernald & Mazzie, 1991). Additionally, CDS is more slowly articulated as opposed to adult-directed speech (Garnica, 1977). Interestingly, infants tend to prefer this speech-style. For example, infants listen longer to speech with these characteristics, especially the pitch characteristics (Fernald, 1985, 1992; Fernald & Kuhl, 1987; Werker & McLeod, 1989; Werker, Pegg, & McLeod, 1994). And, infants respond more to their own mother's voice when speaking 'motherese' (Mehler et al., 1978; Glenn & Cunningham, 1983). However, from a phonetic point of view, in adult-directed speech, in which high and low tones are rapidly alternated and the sequence of sound will split into two perceptually separate groups. By contrast, this is greatly reduced when transitions between successive tones are gradual and continuous as in CDS. Thus, an expanded pitch range (as in CDS) allows greater acoustic contrast among individual elements in utterances. Bregman & Dannenbring (1973) argued that this perceptual integrity of utterances may be enhanced by the use of smooth and continuous pitch excursions. Based on these findings, the question arises what function this speech style has. For example, Kagan (1970) has claimed that exaggerated pitch modulations of child directed speech (CDS) could provide optimal auditory signals for engaging and holding the infant's attention. Additionally, Fernald & Mazzie (1991) suggest that CDS occurs in order to encourage social interaction. And, Fernald, Taeschner et. al. (1989) suggests that this prosodic behavior has a developmental function by facilitating speech processing and language comprehension because prosodic highlighting supports language learning. Thus, it seems as if the speech style that adults use when talking to young infants is strongly related to the acquisition of language. And, as we have seen in the previous chapters of this thesis, children do in fact use the intonational form of an utterance in order to find out its meaning. However, the question remains how children learn to use intonation appropriately. As we have seen in Chapter 4.2., children do have the ability to use intonation for the distinction of the informational status of target referents. However, this ability seems to develop with age as the older children behave more in the adult direction when using intonation for the realization of target referents. This suggests that there is coherence between young children's realization of referents concerning their

informational status within a discourse and the speech they are exposed to in everyday life. Thus, it is an interesting question as to what role the input plays in this development. To my knowledge, there are no studies to date that examine the way in which children's productive use of intonation is influenced by the speech they hear. Thus, in this study, I systematically investigated adult's intonational realization when speaking to children using the same method as in the previous study. Additionally, I compared this study to the results from the previous study, i.e. to the adult-adult realizations as well as to the two child groups.

9.2. Data & Method

In order to find out more about the role of input in young children's intonational development, I asked parents to describe the same picture as in the previous study (cf. Chapter 8). By using exactly the same method as for the two child groups and the adults (talking to adults), it was possible to directly compare the intonational realization of the informational status of target referents from parents talking to their young children with those from the children and adults (talking to adults).

Participants

Eight parents (1 father³⁴, 7 mothers) of 2 year old children (range 2;0 - 2;6, mean= 2;3) were included in the study. Participants were recruited from a database of parents who had volunteered to participate in psychological studies. Two additional fathers were tested but excluded from the study because they did not talk at all to their children (1) or they described the scenes in direct speech (1). Participants came from diverse socioeconomic backgrounds and were from a German medium-sized city. They were raised in the same dialectal environment as the participants from the study presented in the previous Chapter. All participants were tested in a sound-proofed room.

Materials, Design and Procedure

Materials, design and procedure were the same as for the adults in the previous study. Thus, no warm up and practice phase was necessary. Parents were brought into a comfortable room, where they were invited to begin describing the picture books to their children whenever they wanted. Before doing so, parents were asked to put their children on their laps. Unfortunately, it was

³⁴ Due to the few numbers of fathers who participated in this study, it is interesting to know that Davidson & Snow (1996) found that fathers are less talkative in both the number of words as well as the amount of time speaking to children. Additionally, Barton & Tomasello (1994) found that fathers are less communicatively responsive and less conversationally competent, i.e. more communicative breakdowns, fewer successful repairs and shorter conversations.

not possible to elicit a controlled corrective realization of the target referent from the parents in child directed speech. Thus, I concentrated on the new and given realizations of target referents. In case of technical problems or questions, the experimenter was present in the test-room during the test, but did not say anything.

Coding and Reliability

For every picture-book description I again separated those intonational units in which the target referent occurred, following the same criteria as in the previous study. I again checked for in how many cases the target referent was uttered. Target referents occurred in the new condition in 96,42 % of all cases and in the given condition in 100 % of all cases. And, parents described the target referent in the given condition in 100 % of all cases by using a full sentence structure.

The recordings were digitized and annotated with the EMU Speech Database System and annotated using the conventions of GToBI. All realizations of the target referent were coded by the first experimenter. An additional phonetically naïve listener was trained in EMU and GToBI. After this training, he coded 25% of all trials for testing reliability (= complete session of two randomly selected parents). The second coder had no information about the context of the utterances, the condition to which the target referent belonged or the judgments of the other judge. There was perfect agreement with the first rater.

9.3. Results and Discussion

Pitch accent type

I again analyzed the data using a generalized linear mixed model (GLMM). The data from the two children groups and the adults as presented in Chapter 8 was combined with the CDS-data from the present study, using subject and word as random effects and condition and age group as fixed effects (family= binomial, link= logit). I again looked at the mean proportion of times parents used one of the three accent types H*, L*, or deaccentuation when talking to children, in each condition (see Figure 25).

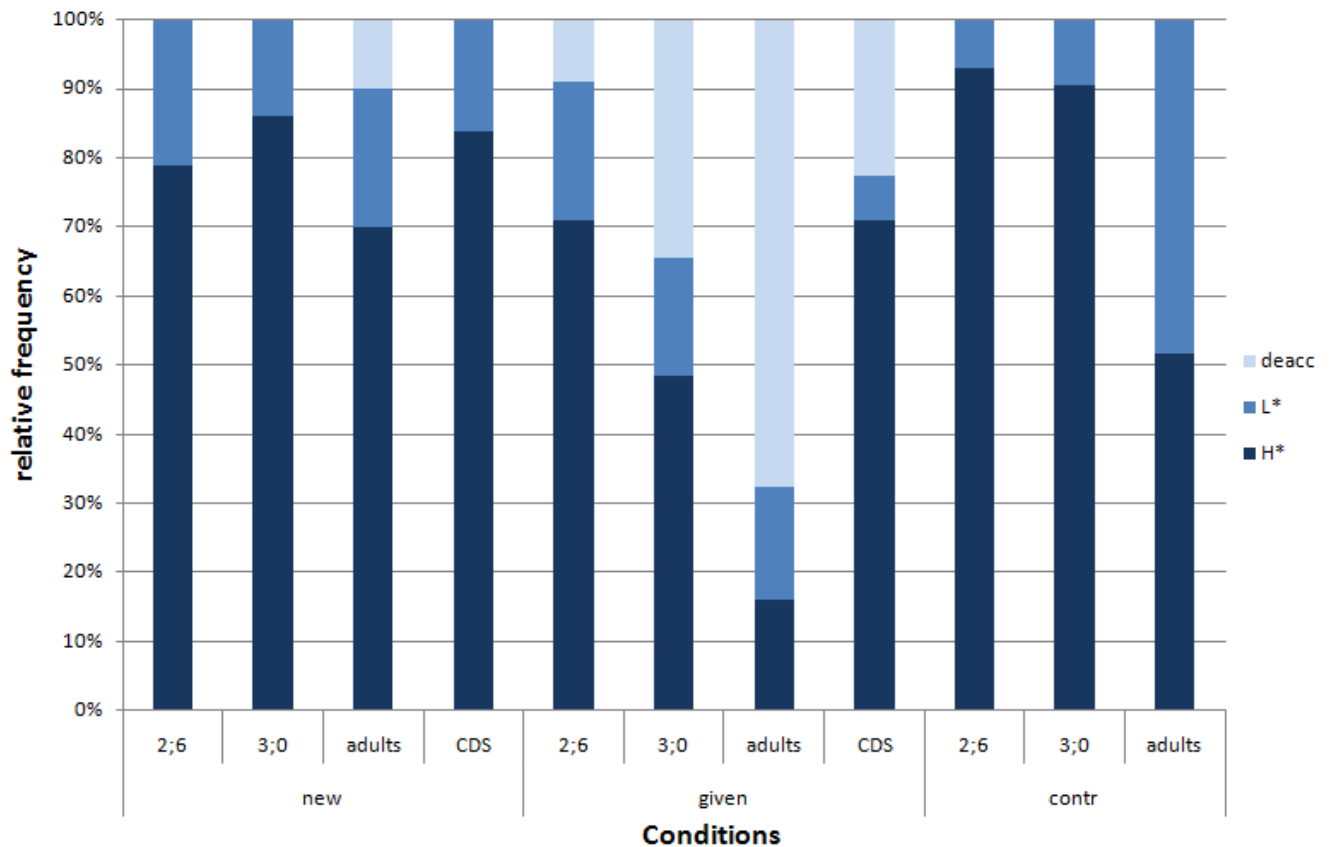


Figure 25: Results of the Pitch Accents with which the target referent was realized in the three conditions. The diagram shows percentages of the use of one of the three accentuation types.

In a first test, I analyzed the use of the H* pitch accent. Statistical analysis of the data revealed that the full model (including condition, group, and their interaction and the random effects) was clearly better than the null model (including only the random effects; likelihood ratio test: $\chi^2=55,6$, $df=5$, $p<0.001$). There was a main effect for age ($\chi^2=16,7$, $df=3$, $p<0.001$) as well as for condition ($\chi^2=39,1$, $df=2$, $p<0.001$). Post-hoc tests revealed no significant difference in parents' use of H* between realizing given and new target referents ($z=0.815$, $p<0.4$), but a significant difference could be found in the CDS data compared to adults use of the high pitched accent when referents were already established ($z=3.424$, $p<0.001$).

The same analysis was made for the use of the low pitched accent type, L*. Statistical analysis of the data set revealed, overall, condition and group significantly explained the accentuation ($\chi^2=12,9$, $df=5$, $p<0.02$) and revealed a main effect for age ($\chi^2=11,3$, $df=3$, $p<0.01$) but not for condition ($\chi^2=0,8$, $df=2$, $p=0.64$). Post-Hoc tests for L* Pitch accent revealed that parents used them significantly less than the older children ($z=-2.066$, $p=0.039$) in the 'new' condition.

For the use of deaccenting the target referents, statistical analysis of the data revealed that condition and group significantly explained the accentuation ($\chi^2=138,23$, $df=5$, $p<0.001$). There was a main effect for age ($\chi^2=29,3$, $df=3$, $p<0.001$) as well as for condition ($\chi^2=111,3$, $df=2$, $p<0.001$).

In a second step, I conducted mixed models as post-hoc tests. The data was one-way error adjusted. A comparison between groups revealed a significant difference for deaccenting the target in the CDS data compared to adults ($z=-3.417$, $p<0.001$).

Pitch range

To analyze the use of pitch range, I again measured the local min and max of the fundamental frequency in Hertz and calculated them into semitones. The data was log-transformed (family = gaussian, link = identity) and tested with a likelihood-ratio test. Analysis of the data revealed, overall, condition and group significantly explained the differences in pitch range ($\chi^2=49.5$, $df=5$, $p<0.001$). I subsequently did a check of assumptions by a visual inspection of residuals plotted against predicted values. The data was then analyzed using a generalized linear mixed model (random effects = subject and word; fixed effects = condition * age). P-values were obtained using Markov-Chain-Monte-Carlo sampling (MCMC). These analyses revealed a significant effect of age (mcmc; $p=0.002$) as well as for condition (mcmc; $p<0.001$). Comparing the data concerning the difference between the conditions 'given' and 'new' revealed a significant difference between the pitch range that parents used to mark the target referent in 'new' and 'given' form when talking to their children (mcmc; $p=0.002$) (see Figure 26).

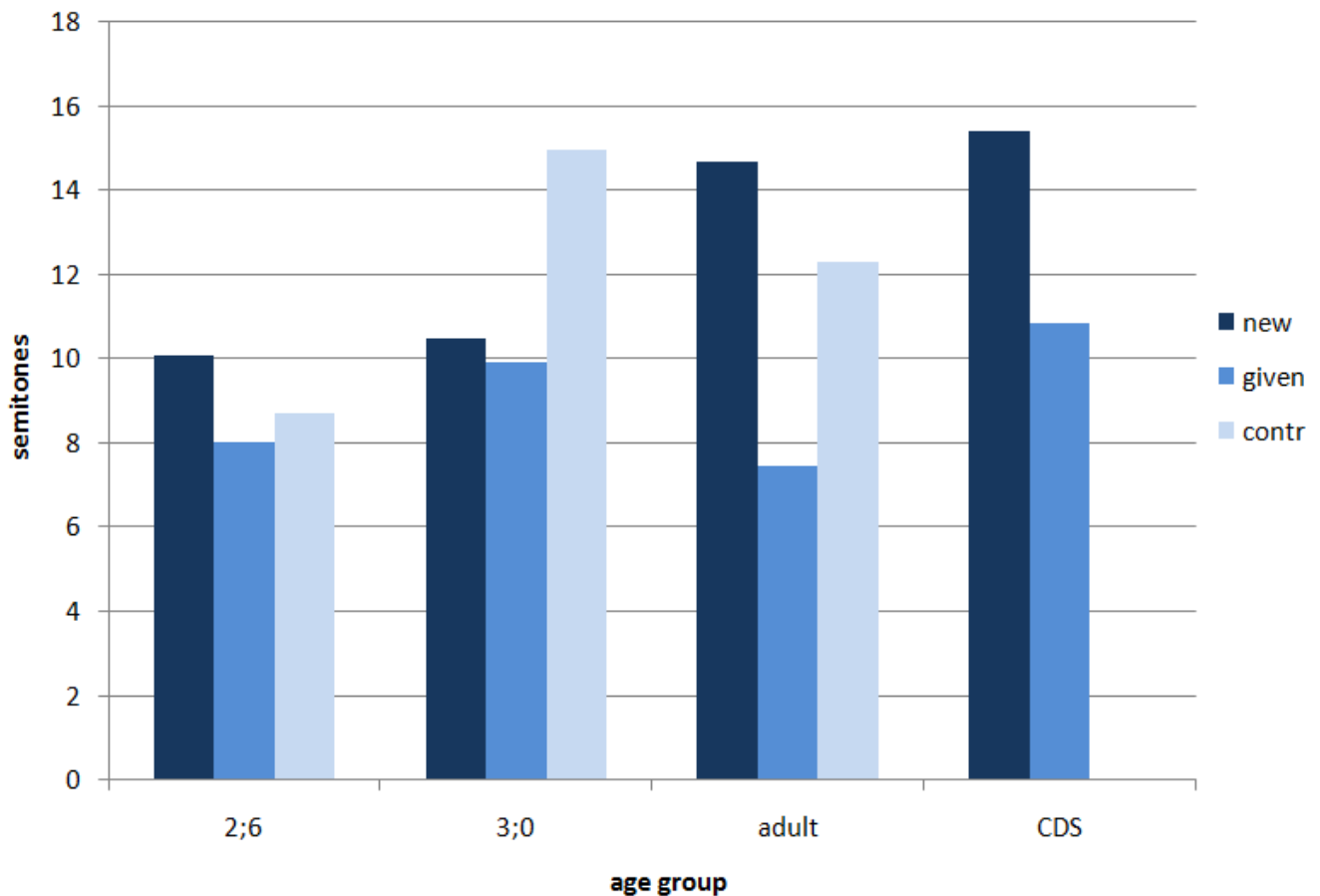


Figure 26: Results of the Pitch Range with which the target referent was realized in the three conditions. The diagram shows the realizations of the target referents in semitones.

The results from this study and the study presented in the previous chapter show that the intonational realizations of target referents that are newly introduced into the discourse are similar in all of the tested groups. But, adults who talk to adults and adults who talk to their young children behave differently in their intonational realizations of target referents that already are established, both in the choice of the pitch accent and in the energy that is put into this realization. The reason for this additional study was to answer the question of why children who are just entering the multi-word stage do not de-accent given information and instead put so much effort into already established information. The answer seems to lie in the speech that is directed to them. Whereas adults (talking to adults) use less high pitched accents and more accentuation to encode given target referents, parents talking to their children behave vice versa – in an identical way to the 2;6 year olds. Thus, it seems plausible that the younger children’s unique intonational behavior in the previous study may come from their copying of adult motherese intonation. The older children have begun to tune into adult intonational patterns when those are speaking to older children and adults.

9.4. General Discussion

Very few studies have looked at young children's intonational realization of referents in discourse, using detailed phonetic and phonological analyses. In the current study, I found that 3-year-old children already make an intonational difference in realizing target referents with different informational statuses in an adult-like way. Thus, children at this age seem to understand that referents already introduced into the discourse are part of the hearer's mental representation. And, they seem to understand that they do not need to make much effort in order to realize that target referent. Instead, they put more effort into the realization of another element in the intonational unit, which may not be part of the common knowledge between the speaker and the hearer. Slightly younger children, however, do not do as older children and adults; i.e. deaccentuate already established target referents. Instead, they use the same high pitched accent for given as for new referents.

This pattern of results could be due to young children's general immaturity in the language learning process. However, it is also possible (and may be a result of this) that young children, in their interaction with adults, hear different accent patterns to older children (to whom adults may use speech that is more like the adult-to-adult speech as the results from the study presented in Chapter 8 suggest). In the second study, therefore, I looked at how adults use intonation to mark the informational status of target referents when speaking to young children, and indeed, the adults displayed the same pattern as the younger children. High pitched accents are a characteristic of the CDS speech register (see Fernald, Taeschner et. al., 1989) and especially F0-variations is a primary acoustic determiner of the infant preference for CDS (Fernald & Kuhl, 1987). This suggests the possibility that the younger children are hearing something different from the older children. In this sense, older children could also be more sensitive to speech around them, e.g. conversations between adults. Both the younger and the older child groups are adapting and learning the use of intonation from the language they hear around them. This view gets supported by findings from Fernald (1985) who could show that the typical CDS pitch contours are perceptual highly salient in the infant's perception. Fernald assumed that this speech style may be particularly well matched to young infant's perceptual and attentional capabilities.

These developmental findings are consistent with those of deRuiter (2010). As already mentioned in Chapter 4.2., she found that German five-year-olds mainly marked new referents with H*, and given referents with deaccentuation (see also Baumann, 2006 and Pierrehumbert & Hirschberg, 1990). However, the children in deRuiter's study also used high pitched accents in nearly 1/4 of all cases. This is consistent with my hypothesis that the use of intonational "norms" is learned. Additionally, this is supported by deRuiter's findings for accessible information. This kind of information normally requires a

more refined control of the speech organs, as the intonational contours are more 'complicated'. For example, due to control over the speech organs, it is easier to realize a H* pitch accent for a referent than a H* !H*. However, children in deRuiter's study realized this type of information similarly to new information, suggesting that they only have a binary distinction of 'active' / 'inactive'. They may have perceived distant referents to be inactive again, leading to a re-activation by the use of accentuation. Taken together, the results from my studies presented in this part of the thesis and those of deRuiter (2010) support the hypothesis that children learn the use of intonation for marking given and new referents from the language they hear and that it takes a considerable period of time to arrive at adult 'norms'.

The remaining question right now is which properties of the intonational distinction develop? First, it seems that the children show a lack of control over the speech organs, which is supported by the findings concerning pitch range. A study done by Chen and Fikkert (2007a) supports this. In their study, two-word utterances of three children at the age of 1;9 – 2;1 years were examined. The authors found that both words in these utterances were accented in most of the cases, regardless of information status. However, the authors claimed that this may not be the whole picture on the phonological marking of focus in two year-olds because "children of this young age are known to have an immature pitch-control system. They may therefore experience difficulty in lowering pitch over the length of a word. This is in fact evidenced by their use of almost complete devoicing to accomplish the effect of deaccenting instead of lowering the pitch" (Chen, in press:8). In contrast, Snow (1998) and Loeb & Allen (2003) found in an imitation task that preschool children did not imitate a rising pattern as accurately as falling pattern in an imitation task. The authors argued that this was due to greater speech production effort when realizing rising patterns as compared to falling patterns. However, although Snow (1998) did examine both imitative and spontaneous speech, the mismatches between the presentation (by the experimenter) and the imitation (from the child) were found in the imitation of yes / no questions (which also Loeb & Allen, 2003 studied). For example, the child should imitate the utterance "Did you take your SOCKS?" Instead of using a rising pattern on the target referent (as presented by the experimenter), the children realized it with a falling pattern. But, as the target referent is already known by both the experimenter and the child in this situation, there is no need to realize the target referent "socks" with a high pitched accent. Instead, the children did use a low-pitch accent, indicating a referent that is given in this situation and thus, absolutely appropriate.

Second, the cognitive abilities seem to play a big role. The appropriate use of intonational pattern within a discourse requires knowledge about the cognitive status of referents within the mind of the listener. Thus, one has to know what others know. And one has to read another's intentions in order to understand communicative goals. This is one crucial point in the acquisition of language, as assumed by the Usage-Based approach. Concerning the differentiation between the informational status of target referents, several approaches e.g., Givón (1990), Vallduví (1992), Lambrecht (1994) are based on

the speaker's assumptions about the cognitive accessibility of referents in the mind of the listener. Chafe (1974, 1976) for example postulates that information can be deaccented when it is already established in the listeners understanding of the context. To do so, the speaker needs to have an understanding about what I know, what you know, what is given and what is new for the other participant(s) of a conversation and so on. The first question thus is what young children really know about the listener's consciousness and the discourse content. Again, several studies have shown that young infants already have this knowledge (cf. Chapter 4.1.), but intonation seems to be a different story. Acquiring the mapping between the cognitive status of target referents within the mind of the listener and the appropriate intonational realization poses an important challenge to (German) children. They not only have to know what others know or do not know, they also need the competence to translate this. This has to be done both in terms of the lexical and syntactic properties of language, but also phonetically. What this means is that children (1) need to have the knowledge about the intonational conventions i.e., how to treat different information, (2) need to control all the physiological properties of the speech organs and (3) have to link all of this to their cognitive knowledge. This view is supported by the results.

To summarize, the results of the two production studies presented in Chapter 8 & 9 show that young children do use intonation to realize the cognitive status of target referents within a discourse. Thus, they understand that there is a difference in the intonational realization of elements within a discourse, depending on their status within the mind of the speaker and the hearer. However, this understanding seems to develop. Between the younger and the older children, a developmental difference in realizing target referents with different informational status was found, converging on adult usage. On the one hand, children seem to learn more about the differentiation between the intonational realization for new and given information from the input. Whereas the younger age group behaves just like parents talking to their children – both in the intonational realization and the energy linked to these realizations – the older children in this study veered away from this. On the other hand, young children have to learn how to control their speech organs and link this to the cognitive understanding about what another person does or does not know. This shows that the acquisition of intonation is an important part in the acquisition of overall cognitive abilities that are needed in order to acquire a language.

10. General discussion

This chapter reviews the major empirical findings of the studies presented in the previous Chapters of this thesis (cf. Chapter 6 - 9). I will discuss how the findings of the current studies relate to general hypotheses and other empirical findings about language development. Finally, I will address open questions, suggest further research, and finish the thesis with a general conclusion.

10.1. Summary and Discussion of empirical findings

The theoretical starting point for the experiments presented in this thesis was the Usage-Based account of language acquisition. As we have seen in Chapter 3.2., this account is based on the assumption that language has cognitive-functional beginnings. The first stipulation is that all representations, from morphemes to words to syntactic constructions, are composed of a form and function. The function as the communicative intention behind a linguistic item or structure (the form) must be formulated in terms of the cognitive structures with which children conceptualize their worlds at different points in development. The question is how intonation fits in into this approach.

In the first study (Chapter 6.1.) I looked at whether young children who have just started the word-learning process use intonational cues in order to find out what another person is referring to. The study was based on previous findings that even the youngest infants can distinguish what is given (and boring) and what is new (and interesting) to another person (e.g. Moll and Tomasello 2007, Tomasello & Akhtar 1995). But, within these studies, children were confronted with multiple cues from which to find a person's referent, including eye-gaze, hand gestures, facial expressions and intonation. In spoken language, the Newness of objects can be clearly distinguished from something given by the use of different pitch accents. For example, a high pitched accent (H*) clearly refers to entities that are newly introduced into the discourse, whereas a low pitched accent and deaccentuation are used for referents that are given. In the current study, I tested whether young word-learners at the age of 20 months are able to take into account these different types of pitch accents when interpreting an utterance. The results suggest that young word-learners use intonation as a way of helping them work out what another person is referring to. This is especially the case when a person is referring to something that is already known. In cases in which a person realized his request for an object with the typical Givenness intonation, the children in my study understood that this intonational form had the function to refer to an old and already known object. However, in order to understand a speaker's intention when referring to a new object, it seems that children need more than just one cue. Thus, I did not find any statistical significance to suggest that 20 months old children understand the request for a new object only based solely on intonation. Rather, in order to gather reliable information about what another person is referring to, it seems that a child needs

a combination of different cues; e.g., body language or additional lexical information. Nevertheless, intonation seems to be a strong cue within this package of cues they can reliably trust on. But, to do so, the function that is conveyed by the intonational form must be supported by another cue.

Related to this, a follow up study was designed with the aim of finding out what role intonation plays in word-learning. To do so, I added Mutual Exclusivity as an additional cue that either supported an already existing label for an object or contradicted it (cf. Chapter 6.2.). The results support the findings from the previous study and suggest that children at the age of 20 months are not exclusively oriented to only one of these conflicting cues but rather to a combination of them.

To summarize, the results of the studies presented in Chapter 6 suggest that children do have an understanding of different types of accent. And, children do use this intonational form in order to find out more about the intention a speaker has. Additionally, it seems that they can use intonation in some sense to learn new words, but only in the absence of more reliable evidence. More importantly, children seem to understand that the intonational form, i.e. the accentuation or deaccentuation of certain words or phrases within an intonation unit reflects a certain function, in this case the reference to an object that is either known or not known. Thus, intonation seems to be an important addition to other cues, not only in word learning but also in the transmission of intentions.

The second study (Chapter 7) builds on the findings of the first study, asking whether the knowledge about the intention conveyed by intonation can pave the way for the comprehension of more complex, syntactic constructions. The question was whether children understand that the intonational realization of an utterance not only has a function when referring to certain objects but also within a more complex linguistic situation. To address this, I examined children's understanding of the basic transitive construction, prototypically used to indicate an agent acting on a patient, as in "The Flomer weefed the Miemel". This kind of construction is of particular importance in language acquisition. Children typically produce spontaneous utterances of this type early on in their language development for the various physical and psychological activities that people perform. To interpret such transitive constructions one needs to understand and to distinguish the different roles of participants in such an event, i.e. to understand the grammatical conventions used to mark the participant roles in the particular language being learned. In most languages the listener has multiple, sometimes redundant cues (e.g., word order, case marking, or animacy) to mark the participants' roles. These cues are acquired step-by-step. For the German language, Dittmar et al. (2008) found that two year olds only understood sentences in which several cues (e.g. case marking and word order) supported each other. At the age of five, children were able to use word order by itself but not case marking, and only 7-year-olds behaved like adults by relying on case marking over word order when these two cues conflicted (e.g. "Den (+accusative) Löwen wieft der (+nominative) Hund" – "The (+accusative) lion is weefing the (+nominative) dog") . However, most studies examining children's understanding of transitive constructions focus on the morpho-syntactic properties of sentences

and ignore the prosodic cue. But, as Weber, Grice & Crocker (2006) demonstrated, adult-listeners use prosodic information in the interpretation of ambiguous SVO and OVS sentences when no clear morphological information is available. Therefore, in my study, I investigated whether five year old German children who were engaged in language learning use prosody for the assignment of participant roles, as has been found for adults. Using a video-pointing task, I embedded transitive OVS utterances in a natural context and presented these utterances as either clearly case marked (e.g. “Den (+accusative) Hund wieft der (+nominative) Hase”) or ambiguous (e.g. “Die (+accusative) Katze wieft die (+nominative) Kuh”). In order to examine the specific role that prosody played for children in resolving the semantic function of the participants, the intonational realization of these constructions was either flat or, to support the syntactic marking of the utterance, characterized by a strong, contrastive pitch accent on the first Nominal phrase.

The results of this study show that children were better at finding the correct agent acting on the correct patient when this was clearly marked by intonation as compared to realizations with no special intonation. And, even when no clear case marking was available, children understood participant roles significantly better when this sentence was realized with the appropriate intonational form rather than when it was presented in a monotonous way. These findings show that children at the age of 5 are able to understand the semantic roles in transitive OVS sentences when appropriate intonation is available. More importantly, in terms of the acquisition of language, they use intonation in order to understand the grammatical conventions of a particular language.

In a follow-up study, where target sentences were presented in a more natural way with a combination of context and intonation, the results were strengthened because the young children used the intonational cue (in combination with case marking and context), as opposed to the competing cue of word order.

In the third study (Chapter 8), I addressed the question of how children, who have just passed the two-word stage of language learning, use intonation in order to realize the cognitive status of target referents within a discourse. For West-Germanic languages like English or German, it is typically assumed that a referent that is accented and realized by a rising contour containing a high pitch accent (H*) introduces new information into the discourse. By contrast, deaccenting, in addition to falling contours containing a low pitch accent (L*), is assumed to refer to already established or given referents (Pierrehumbert & Hirschberg 1990, Baumann 2006). To understand and to realize these linguistic conventions is an essential step. In order to convey information and intentions in the best way, the appropriate intonational form must be chosen. In the current study I investigate whether German learning children between the ages of 2;6 to 3;0 are able to use different types of pitch accents to realize the informational status of target referents within semi-spontaneous speech. Using a story telling task, I designed picture books in which a target referent was either new or given within the discourse. I then analyzed the data measuring the kind of pitch accent (H*, L* or deaccentuation) with which the target referent was realized.

Additionally, these results were compared with the results from an adult control group. Whereas the results for this control group are similar to those found by Baumann (2006) (adults accented new information and deaccented given information) the findings for both child groups differ. Unlike the findings for adults, I found that children at the age of 2;6 and 3;0 years tended to realize both new and given information with an high pitched accent. Moreover, I found a development in children's intonational realization of the informational status of target referents. Thus, the 2;6 year old children realized the target referent in the given condition significantly more often with a H* pitch accent than the 3 year olds, who deaccented the already established referent significantly more often than the younger ones.

Based on these findings and the question of why the younger children do not deaccent given information, I hypothesized that this could be due to the speech to which young children are exposed to in everyday life. The accenting of information, even if it is given information, is a characteristic of the motherese speech register used by most western, middle class parents (e.g. Fischer & Tokura, 1995). From an acoustic point of view, motherese has a clear signature (high pitch, exaggerated intonation contours) and has been shown to be preferred by infants over adult-directed speech and might assist infants during the language acquisition process (Kuhl 2004). In order to address this question, I used the same method as in the previous study and analyzed the intonational form that parents use when talking to their 2 year old children (cf. Chapter 9). When compared to the results from the first part of the study, I found that, as with the younger age group, parents do not differ in their use of H* between given and new.

To summarize, the two studies presented in Part III suggest a development in children's intonational realization of the informational status of target referents. Furthermore, when parents talk to their young children they behave differently to the way that adults talk to other adults. Instead of deaccenting already established referents, parents treat these as if they were new. Interestingly, children seem to adopt this behavior. Whereas the younger age group realized given target referents in a way that was similar to how their parents had presented them, the older children shifted more towards adults' non-CDS behavior. This suggests that encoding the informational status of target referents by intonation develops with experience.

Taken together, the studies presented in this thesis have raised three major issues. First, I argue that the results of the studies presented in this thesis show that the development of intonational behavior (both in production and comprehension) is strongly related to the overall pragmatic and social-cognitive abilities that children need in order to acquire a language. In this sense, intonation is an important part in understanding another's communicative intention and fits perfectly into the Usage-Based approach to language acquisition. Within this approach, it is assumed that language consists of constructions. Children are exposed to language all the time, and this input consists of a 'language package'. This package includes all kinds of information, e.g. morphological marker, lexical referents, grammatical constructions and

intonation. The child has the task to pull this package apart and to sort out the different kinds of information that is provided by the input. Within this package of information, intonation has a special function, as it can be independent of the syntactic structure. As we have seen, the sentence "The boy has a red jacket" can be uttered in different ways. Depending on the importance of certain parts, a speaker can mark them by accentuation. Thus, if it is especially important that it is a boy (as opposed to a girl), a speaker would say "The BOY has a red jacket". When the colour of the jacket is of special importance, this leads to a realization like "The boy has a RED jacket", and so on. However, the bigger point is that the child has to understand that the form of the intonational realization has a pragmatic function within the message. A language learning child has to 'unpack' the information she gets and find out the specific role of intonation within the package that is provided by the input. Thus, the development of both production and comprehension of the pragmatic and social-cognitive functions of intonation is strongly related to the overall cognitive abilities that are needed to learn, and to understand, the intentional aspects of human communication.

Second, my studies have shown that language learning children do use the intonational form of an utterance from early on in order to understand another's intention. Young language learning children do understand that a certain intonational form (the accentuation of certain words or parts of an utterance) has a function within the message the speaker is conveying (i.e. the particular importance of this part within the utterance). However, I found that initially these comprehensions studies are only relatively independent of other cues in the message. Children also seem to use a certain intonational form (once they understood what effect this form has) in order to convey their own communicative intentions. As the two production studies in this thesis suggest is this usage a developmental one. It is not clear what exactly it is that develops. The question about this development leads us to the third and maybe most important issue for understanding the development of young children's use of intonation. The studies presented in this thesis suggest that children seem to be faced with several problems. Three factors seem to influence young children's development in realizing the intonational form of an utterance. First, children need to acquire knowledge about the intonational conventions of the language they are 'growing into' as the studies show that this is developmental. This is not surprising because cross-linguistic differences mean, for example children learning Chinese as their first language have to understand that pitch variations result in morpho-lexical differences, whereas a child, growing up in a West-Germanic language environment mainly needs pitch variation for postlexical distinction such as to mark the informational status of target referents. And, more important for this thesis, a child has to come to understand the intentional aspects of a situation, not only how a situation is described, but also why and how this is reflected in how it is said. For example, the German learning child has to understand the function behind an intonational form. A learner who knows about the existence of these functions will not only learn to express them, but will also use them to

interpret language he hears in a more analytic way, thus reducing the danger of attributing unexpected intonation patterns as (solely) a function of the attitude or emotional state of the speaker. Second, in order to convey information in the best way, children have to understand what other people know and what they do not know. For example, once they understand that people are more interested and more likely to be excited about new things than about "old news", they can use this knowledge for the interpretation of other people's behavior. Third, children need the ability to link this knowledge to the physiological properties of their speech organs. However, as we have seen in the introductory Chapter (Chapter 2.3.2.), some of this seems to be instinctive. Certain biological devices, for example fear, anger, happiness, manifest themselves in particular bodily behaviors – the vocalization related to this emotion automatically assimilates to these bodily expressions. For example, in the case of surprise, the blood pressure increases as does our rate of breathing. This leads to more air in the lungs which in turn results in the accentuation in speech. In the event of something unexpected or special happening, therefore, the emotional state activated by this produces a certain vocalization. This means, in my opinion, that the linguistic use of intonational patterns (e.g. the distinction between new and given information) is strongly related to its paralinguistic use, i.e. its affective meaning. This affective meaning seems to be directly derived from the speaker's emotional state at the moment of that vocalization. Thus, the meaning of an intonational contour can be directly derived from the underlying biological properties. For example, a speaker who is very glad and excited about something will automatically encode this excitement in his utterance. He will speak louder and with an exaggerated intonational contour as depicted by Gussenhoven's "Effort Code" (cf. Chapter 2.3.2.).

To summarize, we have seen that intonation can be realized both purposefully and accidentally. In the latter case, biological devices seem to be responsible for indicating dominance, fear or happiness. This, on the other hand, could have developed for linguistic purposes. It seems plausible that the grammatical use of intonation e.g., marking given and new information, is strongly related to intonational universals. For example, people tend to be excited about new things, excitement results in certain bodily expressions e.g., hand gestures, pointing, faster breathing, more air in the lungs, accentuation and so on. As they try to talk about new things, bodily expressions become part of the intentional message.

10.2. Open Questions and Future Research

The studies presented in this thesis indicate that German learning children understand the intentions reflected by the use of intonation. However, since this is a very complex issue, the data from the current studies cannot

answer completely the exact manner in which this understanding develops. Thus, further research is necessary.

In order to understand more about the referential function of intonation, it would be necessary to distinguish the different cues that children rely on. For example, what role does intonation play in combination with each of the other cues, e.g. hand gestures, facial expressions, eye-gaze and words. And, what happens when these cues are put into conflict. For example, Grassmann & Tomasello (2010) showed that children at the age of 2 & 4 years rely most heavily on pragmatic information (e.g. in a pointing gesture), and only secondarily on lexical conventions and principles. The study presented in Chapter 6 shows that Mutual exclusivity is a very strong cue (maybe the strongest) for young-word learning children. The question about the reliability of other cues arises and how they interact with intonation. For example, what role does intonation play when it co-occurs together with pointing or eye-gaze? And, what happens when these cues contradict each other?

The second study that was presented in this thesis made a huge step (from word learning in 20 months old children to the understanding of grammatical construction in 5 year old children). The question at this point is how far the children have come with unpacking intonation from the overall input. Do they understand that the intonational realization of utterances have a certain function? In order to find out more about the role of intonation in grammatical constructions, it would be necessary to do research in this area with younger children who are only just beginning to be exposed to grammatical constructions, e.g. intransitive constructions. Additionally, it would be useful to test children's understanding in more complex grammatical constructions, e.g. in combination with relative clauses. And, to find out more about their competence in this area, production studies would be of especially importance.

In case of the production studies, as presented in Part III of this thesis, there has been hardly any previous work on the intonational realization of very young children. Although there is clear scope for detailed further research, it was sufficient for this initial study to subsume the range of possible intonational contours into three classes of pitch accent types, namely H*, L* and deaccentuation. In future, in order to find out more about the development of the control over the speech organs, a more sophisticated analysis seems to be necessary. Additionally, a narrower investigation of the interaction with syntactic structure seems of importance, because word order variations, for instance SVO and OVS sentences, used in order to describe the stories in the picture books, may have intonational consequences. Related to this, further analyses of the placement of pitch accents (nuclear / prenuclear) would be of particular importance.

Overall, the present studies examined German learning children, and an obvious next step would be to extend the findings to research within other languages. Cross linguistic comparisons of the acquisition of languages that differ in their prosodic structure are necessary and important. For example, in stress-accented languages like German, English or Dutch, accentuation is mainly used for the marking of informational status at the level of utterance. In Pitch accent

languages like Swedish or Norwegian, children are additionally faced with the task of distinguishing a number of words based only on word stress. Swedish and Norwegian differentiate between two kinds of accents, often referred to as 'Accent 1' and 'Accent 2' (e.g. Öhman, 1967; Gårding & Lindblad, 1973). For example, the word 'anden' has two meanings: 'duck' and 'spirit'. Which of the two meanings is intended depends on the intonational realization (see Bruce, 1977). Finally, tonal languages, for example Yucatec Maya, use pitch variation and tonal contrasts for lexical and morphological marking in order to make distinctions at word level. What this means is that children who grow up in different prosodic language systems have to master many different tasks regarding the acquisition of prosody. A cross-linguistic comparison of these languages would give deeper insight into how children acquire intonation and would help to understand the acquisition of language and the role of intonation within this process as a whole.

Furthermore, the studies in this thesis deal with children who clearly have passed the preverbal stage. Thus, it is an interesting question whether pre-verbal infants use prosody in order to understand others' intentions. As we have seen in Chapter 4.1., infants show some prelinguistic abilities that they use in order to influence the psychological states of others. For example, infants point to an interesting event when the adult had not yet seen it (Liszkowski et al., 2007b) and to inform an adult about the location of an object when he is looking for that object (Liszkowski et al., 2006). In the same way as pointing seems to be a natural way to inform others and thus to change their mental state, this job can be done with intonation as well ("pointing with words" – so to speak). However, it is unclear whether prelinguistic pointing is combined with a certain prosodic behavior, in order to strengthen the pointing gesture. Further research into the relationship between intonation and pointing in preverbal infants would certainly be of great interest.

In relation to this, and to understand more about the evolutionary aspects of intonation, it seems necessary to find out more about the relation between the paralinguistic meaning of intonation and its development towards becoming linguistic conventions. An interesting scenario would be the examination of young children's comprehension and production of different emotional states in order to understand another's intentions.

10.3. Principal Conclusions

Children use a variety of social and general cognitive skills in order to understand the world around them. In this sense, the acquisition of language requires a certain mind-reading ability. The use of a particular intonational pattern mirrors the speaker's knowledge and what the speaker thinks about the hearer's knowledge. Thus, intonation is an important instrument for young children in order to understand what another person refers to or what that person has in

mind – the prerequisite for understanding how the world around them works. More importantly regarding this thesis, intonation is a prerequisite for the acquisition of language from an early age. Despite a number of open questions that need to be addressed in future work, the studies presented in this thesis show that young children are able to understand a speaker's communicative intention based on intonation.

The current studies are just a first step towards fully understanding children's use of prosody, in particular intonation, in the language acquisition process. It is likely that prosody interacts in complex ways with a number of different grammatical and pragmatic properties of language. This interplay between lexical, grammatical, and prosodic properties for a particular language must be learned. Ultimately, in order to understand the process of language acquisition, the role of intonation must be taken into account.

11. References

- Abbot-Smith, K. & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *Linguistic Review*, 23(3), 275-290.
- Akhtar, N., & Tomasello, M. (1996). Two-year-olds learn words for absent objects and actions. *British Journal of Developmental Psychology*, 14(Pt 1), 79-93.
- Akhtar, N. & Tomasello, M. (1997). Young children's productivity with word order and word morphology. *Developmental Psychology*, 33, 952-965.
- Akhtar, N., Carpenter, M. & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, 67(2), 635-645.
- Allbritton D. W., McKoon, G. & Ratcliff R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22, 714-735.
- Allen, S. E. M. (1996). *Aspects of Argument Structure Acquisition in Inuktitut*. Amsterdam: Benjamins.
- Arnold, J. E. (2008). "THE BACON" Not "the Bacon": How Children and Adults Understand Accented and Unaccented Noun Phrases, *Cognition* 108, 69-99.
- Atkinson, M. (1992). *Children's syntax. An introduction to principles and parameters theory*. Oxford: Blackwell.
- Baayen, R.H. (2008). *Analyzing Linguistic Data*. Cambridge University Press. Cambridge.
- Bach, K. & Harnish R. M. (1979), *Linguistic Communication and Speech Acts*, Cambridge, Mass.: MIT Press.
- Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, 55, 1278-1289.
- Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development*. Hillsdale, NJ: Erlbaum.

- Baldwin, D.A. & Moses, L.J. (2001). Links between social understanding and early word learning: Challenges to current accounts. *Social Development*, 10, 309-329.
- Baltaxe, C. (1984). The use of contrastive stress in normal, aphasic and autistic children. *Journal of Speech and Hearing Research*, 27, 97-105.
- Bannard, C., Lieven, E. & Tomasello, M. (2009). Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences* 106, 17284-17289.
- Tomasello, M. & Barton, M. (1994). Learning words in non-ostensive contexts. *Developmental Psychology*, 30, 639-650.
- Bates, E. (1979). *The Emergence of Symbols*. New York: Academic Press.
- Bates, E., MacWhinney, B., Caselli, C., Devescovi A., Natale, F. & Venza, V. (1984). A cross-linguistic study of the development of sentence interpretation strategies. *Child Development* 55, 341–354.
- Bates, E. & MacWhinney B., (1987). Competition, variation, and. language learning. In MacWhinney, Brian (ed.), *Mechanisms of Language Acquisition*. Hillsdale, NJ: Erlbaum, 157–193.
- Bates, E. & MacWhinney, B. (1989). Functionalism and the competition model. In MacWhinney, B. and E. Bates (eds.), *The Crosslinguistic Study of Sentence Processing*. New York: Cambridge University Press, 3–76.
- Bates, DM, Maechler, M. & Dai, B. (2008). lme4: Linear mixed-effects models using S4 classes. *R package version .999375-24*.
- Baumann, St. (2006). *The Intonation of Givenness - Evidence from German*. Linguistische Arbeiten 508, Tübingen: Niemeyer.
- Baumann, St. & Hadelich, K. (2003). Accent type and givenness: an experiment with auditory and visual priming. In: *Proceedings of the 15th ICPhS Barcelona*.1811–1814.
- Baumann, St. & Grice, M. (2006). The Intonation of Accessibility. *Journal of Pragmatics* 38 (10), 1636-1657.
- Baumann, Stefan, Doris Mücke & Johannes Becker (2010). Expression of Second Occurrence Focus in German. *Linguistische Berichte* 221. 61-78.
- Beckman, M. & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston & M. Beckman (eds.), *Papers in Laboratory Phonology I*, 179-200. Cambridge, UK: Cambridge University Press.

- Behrens, H. (2000): Rezension von Steven Gillis und Annick de Houwer (Eds):
The Acquisition of Dutch. Amsterdam: Benjamins. *Journal of Child Language*, 27, 437- 442.
- Bernstein-Ratner, N. (1985). Dissociations between Vowel Durations and Formant Frequency Characteristics. *Journal of Speech and Hearing Research*, 28, 255-264.
- van Bezooijen, R. (1984) *The characteristics and recognizability of vocal expression of emotions*. Dordrecht. The Netherlands: Foris.
- Bever, T.G., Fodor, J.A., & Weksel, W. (1971). Theoretical notes on the acquisition of syntax: A critique of 'contextual generalization'. A. Baradon and W.F. Leopold, *Child language: A book of readings*, Prentice-Hall, Englewood Cliffs, NJ, 263–278.
- Biemans, M. (2000). *Gender variation in voice quality*. PhD thesis, Katholieke Universiteit Nijmegen.
- Borden, G. J. & Harris, K.S. (1984). *Speech Science Primer: Physiology, Acoustics and Perception of Speech* (2nd edition). Baltimore: Williams & Wilkins.
- Bowerman, M. (1973). *Early syntactic development*. London: C.U.P
- Braine, M. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development* 41 (1).
- Bregman, A.S. & Dannenbring, G. (1973). The effect of continuity on auditory stream segregation. *Perception & Psychophysics*, 13, 308-312.
- Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.
- Bruce, G. (1977). *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.
- Bybee, J. (2006). From usage to grammar: the mind's response to repetition. *Language*, 82(4), 529-551.
- Carlson, K., Frazier, L. & Clifton, C. (2009). How prosody constrains comprehension: A limited effect of prosodic packaging. *Lingua* 119, 1066–1082.
- Carpenter, M., Nagell, T., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63(4, Serial No. 255).

- Cassidy, S. & Harrington, J. (2001). Multi-level annotation in the Emu speech database management system. *Speech Communication*, 33, 61-77.
- Chafe, W. (1974). Language and Consciousness. *Language* 50, 111-133.
- Chafe, W. (1976). Givenness, Contractiveness, Definiteness, Subjects, Topics and Point of View. In: Charles Li, (ed.). *Subject and Topic*. New York: Academic Press, 25-56.
- Chafe, W. (1994). *Discourse, Consciousness, and Time*. Chicago/London: University of Chicago Press.
- Chan, A., Lieven, E. & Tomasello, M. (2009) Children's understanding of the agent-patient relations in the transitive construction: Cross-linguistic comparisons between Cantonese, German, and English, *Cognitive Linguistics* 20 (2), 267–300.
- Chen, A. (2007). Intonational realisation of topic and focus by Dutch-acquiring 4- to 5-year-olds. In J. Trouvain, & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, 1553-1556.
- Chen, A. (in press). The developmental path of phonological encoding of focus in Dutch. In: S. Frota, P. Prieto, and G. Elordieta (eds.) *Prosodic production, perception and comprehension*. Springer Verlag.
- Chen, A., & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. In: J. Trouvain, & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, 315-320.
- Choi, Y. & Mazuka, R. (2003). Young children's use of prosody in sentence parsing. *Journal of Psycholinguistic Research*. 32, 197-217.
- Chomsky, N. (1959). A review of B. F. Skinner's Verbal behavior. *Language* 35, 26– 58.
- Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.
- Chomsky, N. (1993). On the nature, use, and acquisition of language. In: A. I. Goldman (Ed.), *Readings in philosophy and cognitive science*, Cambridge, MA: The MIT Press, 511-534
- Chomsky, N. (1999). Derivation by phase. *MIT Occasional Papers in Linguistics* 18, Cambridge, MA: MIT Linguistics Department.
- Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

- Corkum, V. & Moore, C. (1995). Development of joint visual attention in infants. In Moore, C. & Dunham, P. J. (Eds.), *Joint attention: Its origins and role in development*. Hillsdale, NJ: Erlbaum, pp. 61-83.
- Couper-Kuhlen, E & Selting, M. (1996) *Prosody in Conversation*. Cambridge: Cambridge UP.
- Crain, S., & Pietroski, P. (2001). Nature, Nurture and Universal Grammar. *Linguistics and Philosophy*, 24, 139-186.
- Cruttenden, A., (1986). *Intonation*. Cambridge: Cambridge University Press.
- Cruttenden, A. (2006). The De-accenting of Given Information: a Cognitive Universal? In: Bernini, G. & M. L. Schwartz (eds.), *Pragmatic Organization of Discourse in the Language of Europe*. The Hague: Mouton de Gruyter. 311-355.
- Crystal, D. (1979). Prosodic development. In: Fletcher, P.J. & Garman, M.A. (eds.). *Language acquisition* (Cambridge: CUP), 33-48, 2nd edn, 1986, 174-97.
- Crystal, D. (1987). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua*, 92, 81–104.
- Cutler, A. & Swinney, D. (1987). Prosody and the Development of Comprehension. *Journal of Child Language*, 14, (1), 145-167.
- Dąbrowska, E. & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics* 16, 437-474.
- Dahan, D. & Bernard, J.M. (1996). Interspeaker variability in emphatic accent production in French. *Language and Speech* 39 (4), 341-374.
- Davidson, R. & Snow, C.E. (1996). Five-year-olds' interactions with fathers versus mothers. *First Language*, 16, 223-242.
- Delin, J. (1995). Presupposition and shared knowledge in *it*-clefts, *Language and Cognitive Processes* 10, 97–120.
- Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, 37, 630–641.
- Diesendruck, G., Markson, L., Akhtar, N., Reudor, A. (2004). Two-year-olds' sensitivity to speakers' intent: An alternative account of Samuelson and Smith. *Developmental Science*, 7, 33–41.

- Dittmar, M., Abbott-Smith, K., Lieven, E. & Tomasello, M. (2008). German Children's Comprehension of Word Order and Case Marking in Causative Sentences, *Child Development* 79 (4), 1152 – 1167.
- Dobson, A. J. (2002). *An Introduction to Generalized Linear Models. Texts in statistical science*. Boca Raton, FL: Chapman & Hall/CRC.
- Durieux, G. & Gillis, St. (2001). Predicting grammatical classes from phonological cues: An empirical test. In: Jürgen Weissenborn & Barbara Höhle (eds.): *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition*, 189–229. Amsterdam: John Benjamins.
- Ekman, P. (1984). Expression and the nature of emotion. In: K. Scherer & P. Ekman (eds). *Approaches to Emotion*. Hillsdale, NJ: Erlbaum, pp. 319-344.
- Ekman, P. (1999). Basic Emotions. In: T. Dalgleish and M. Power (Eds.). *Handbook of Cognition and Emotion*. Sussex, U.K.: John Wiley & Sons, Ltd.
- Elman, J. L., Bates, E. A. Johnson, M. H., Karmiloff-Smith, A., Parisi, D.& Plunkett, K. (1996). *Rethinking Innateness. A Connectionist Perspective on Development*. Cambridge, MA: The MIT Press.
- Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, 8, 181-195.
- Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In Papousek, H., Jurgens, U., & Papousek, M. (Eds.), *Nonverbal vocal communication: Comparative and developmental approaches*. Cambridge: Cambridge University Press, 262-282
- Fernald, A. & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20(1), 104-113.
- Fernald, A. & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, 10, 279-293.
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B.& Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language* 16, 477–501.
- Fernald, A. & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology* 27, 209–221.

- Fisher, C., & Tokura H. (1995). The given-new contract in speech to infants. *Journal of Memory and Language*, 34, 287–31.
- Fry, D.B. (1955). Duration and Intensity as Physical Correlates of Linguistic Stress. *Journal of the Acoustical Society of America* 27, 765-768.
- Fry, D.B. (1958). Experiments in the Perception of Stress. *Language and Speech* 1, 126-152.
- Garnica, O. (1977). Some prosodic and paralinguistic features of speech to young children. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition*. Cambridge, England: Cambridge University Press.
- Gårding, E. & Lindblad, P. (1973). Constancy and variation in Swedish word accent patterns. *Lund Working Papers* 7:36–110.
- Gathercole, V. C. (1989). Contrast: A semantic constraint? *Journal of Child Language*, 16, 685–702.
- Gathercole, V., Mueller C., Eugenia S., & Pilar., S. (1999). The early acquisition of Spanish verbal morphology: Across-the-board or piecemeal knowledge? *International Journal of Bilingualism* 3 (2 & 3), 133-182.
- Gentner, D., & Namy, L. L. (2006). Analogical Processes in Language Learning. *Current Directions in Psychological Science*, 15 (6), 297-301.
- Gerken, L. (1996). Phonological and distributional information in syntax acquisition. In: James L. Morgan & Katherine Demuth (eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, Mahwah, NJ: Lawrence Erlbaum, 411–425.
- Givón, T. (1990). *Syntax: A Functional-Typological Introduction*, Vol. II. Amsterdam and Philadelphia: John Benjamins.
- Gleitman, L. (1990). The structural sources of verb meaning. *Language Acquisition*, 1, 3-55.
- Gleitman, L. & Wanner, E. (1982). Language acquisition: The state of the state of the art. In Eric Wanner & Lila R. Gleitman (eds.), *Language acquisition: The state of the art*, Cambridge, MA: Cambridge University Press, 3–48.
- Goldberg, A. E. (1995). *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- Goldberg, A. E. (2006). *Constructions at work: the nature of generalizations in language*. Oxford: Oxford University Press.

- Gomez, R. & Gerken, L. (1999). Artificial grammar learning by 1-year-olds lead to specific and abstract knowledge. *Cognition* 70, 109-135.
- Grassmann, S. & Tomasello, M. (2007). Two-year-olds use primary sentence accent to learn new words. *Journal of Child Language*, 34, 677-687.
- Grassmann & Tomasello (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science* 13:1, 252-263.
- Grice, H.P. (1975). Logic and Conversation. In: D. Davidson and G. Harman (eds.). *The logic of grammar*. Encino, California: Dickenson, 64-75.
- Grice, Martine (2006). Intonation, In: K. Brown (ed.). *Encyclopedia of Language and Linguistics, 2nd Edition*, Elsevier: Oxford, Vol 5, 778-788.
- Grice, M., Reyelt, M., Benzmüller, R., Mayer, J. & Batliner, A. (1996). Consistency in Transcription and Labelling of German Intonation with GToBI, *Proc Fourth International Conference on Spoken Language Processing*, Philadelphia, 1716-1719.
- Grice, M. & Baumann, St. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte* 191, 267-298.
- Grice, M., Baumann, St. & Benzmüller, R. (2005). German Intonation in Autosegmental-Metrical Phonology. In: Sun-Ah Jun (ed.), *Prosodic Typology. The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. 55-83
- Grice, M. & Baumann, St. (2007). An Introduction to Intonation – Functions and Models. In: Trouvain, Jürgen & Ulrike Gut (eds.): *Non-Native Prosody. Phonetic Description and Teaching Practice*. Berlin, New York: De Gruyter (= Trends in Linguistics. Studies and Monographs [TiLSM] 186). 25-51.
- Grosse, G., Behne, T., Carpenter, M. & Tomasello, M. (in press). Infants communicate in order to be understood. *Developmental Psychology*.
- Goldsmith, J. A. (1976). An Overview of Autosegmental Phonology. *Linguistic Analysis* 2, 23-68.
- Guasti, M. T., Christophe, A., van Ooyen, B. & Nespors, M. (2001). Prelexical setting of the head complement parameter. In Jürgen Weissenborn & Barbara Höhle (eds.), *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition* 1, Amsterdam: John Benjamins, 231–248.

- Glenn, S. M., & Cunningham, C. C. (1982). Recognition of the familiar words of nursery rhymes by handicapped and nonhandicapped infants. *Journal of Child Psychology and Child Psychiatry*, 23, 3 19-327.
- Grimm, H. (2001). *Sprachentwicklungstest für drei- bis fünf- jährige Kinder. Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen*. Göttingen, Germany: Hogrefe.
- Gundel, J. & Fretheim T. (2004). Topic and Focus. In: L.R.Horn & G. Ward (Eds.) *The Handbook of pragmatics*. Malden, MA:Blackwell, 175-196.
- Gundel, J. K., Hedberg, N. & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language* 69, 274-307.
- Gussenhoven, C. (1983). Focus, Mode, and the Nucleus. *Journal of Linguistics* 19, 377-417.
- Gussenhoven, C. (1984). On the grammar and semantics of sentence accents. Dordrecht: Foris.
- Gussenhoven, C. (2002). Intonation and Interpretation: Phonetics and Phonology. *Proceedings 1st Int. Conference on Speech Prosody*, Aix-en-Provence, 47-57.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Gussenhoven, C. (2005). Semantics of prosody. In: Brown, K. (ed.). *Encyclopedia of Language and Linguistics*, 2nd volume. Oxford: Elsevier. Volume 11, article 4319, 170-173.
- Halliday, M.A.K. (1967b). Notes on Transitivity and Theme in English, Part 2, *Journal of Linguistics* 3, 199-244.
- Hart, J. 't, Collier, R. & Cohen, A. (1990). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach*. Cambridge: Cambridge University Press.
- Hauser, M. D., Chomsky, N. & Fitch, W. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298. 1569–1579.
- Hayes, B. (1982). Extrametricality and English Stress. *Linguistic Inquiry* 13, 227-276.
- Hermes, A., Becker, J., Mücke, D., Baumann, St. & Grice, M. (2008). Articulatory Gestures and Focus Marking in German. *Proceedings of the 4th Conference on Speech Prosody 2008*, Campinas, Brasil. 457-460.

- Hirsh-Pasek, K., & Golinkoff, R. M. (1996). *The Origins of grammar: Evidence from comprehension*. Cambridge, MA: The MIT Press.
- Hirsh-Pasek, K., Kemler Nelson, D.G., Jusczyk, P.W., Wright, K., Cassidy, B.D. & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition* 26. 269–286.
- Hornby, P.A. & Hass, W.A. (1970): *Use of contrastive stress by preschool children*, *Journal of Speech and Hearing Research* 13, 395-399.
- Höhle, B. (2009). Bootstrapping mechanisms in first language acquisition. *Linguistics*, 47 (2), 359-382.
- Höhle, B. & Weissenborn. J. (2000). The origins of syntactic knowledge: Recognition of determiners in one year old German children. In S. Catherine Howell, Sarah A. Fish & Thea Keith-Lucas (eds.), *Proceedings of the 24th annual Boston University conference on language development*, Somerville: Cascadilla Press, 418–429.
- Höhle, B. & Weissenborn. J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science* 6, 122–127.
- Höhle, B. & Weissenborn. J., Kiefer, D., Schulz, A. & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy* 5, 341–353.
- Ibbotson, P. & Tomasello, M. (2009) Prototype constructions in early language acquisition, *Language and Cognition*, 1 (1), 59–85,
- Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.
- Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D.G., Kennedy, L.G., Woodward, A. & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology* 24, 252–293.
- Jusczyk, P.W., Houston D.M. & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology* 39, 159–207.
- Kagan, J. (1970). Attention and psychological change in the young child. Analysis of early determinants of attention provides insights into the nature of psychological growth. *Science*, 170, 826-832.

- Kingston, J. (1991). Integrating Articulations in the Perception of Vowel Height. *Phonetica* 48, 149-179.
- Kohler, K. (1991a). A model of German intonation. *AIPUK (Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung, Universität Kiel)* 25, 295–360.
- Kohler, K. (1995). *Einführung in die Phonetik des Deutschen*. (Grundlagen der Germanistik 20). Berlin: Schmidt.
- Kuhl, P.K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience* 5, 831–843.
- Ladd, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.
- Ladd, D. R. & Silverman, K. (1984). Vowel Intrinsic Pitch in Connected Speech. *Phonetica* 41, 31-40.
- Ladefoged, P. (1962). *Elements of Acoustic Phonetics*. Chicago: University of Chicago Press.
- Lambrecht, K. (1994). *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar (Vol. 2)*. Stanford: Stanford University Press.
- Langacker, R. W. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language*. Stanford: CSLI Publications, 1-63.
- Lehiste, I. & Peterson, G.E. (1961). Some Basic Considerations in the Analysis of Intonation. *Journal of the Acoustical Society of America* 33, 419-425.
- Lewis, M.M. (1951). *Infant Speech*, London: Routledge.
- Lieberman, P. (1967). *Intonation, Perception, and Language*. Cambridge, MA: MIT Press.
- Liberman, M. (1975) [1979]. *The Intonational System of English*. New York: Garland.
- Liberman, M. & Prince, A. (1977). On Stress and Linguistic Rhythm. *Linguistic Inquiry* 8, 249-336.
- Lieven, E., Pine, J., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24(1), 187-219.

- Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language*, 30 (2), 333–367.
- Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science*, 7(3), 297-307.
- Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). Twelve- and 18-month-olds point to provide information for others. *Journal of Cognition and Development*, 7, 173-187.
- Liszkowski, U., Carpenter, M., & Tomasello, M. (2007a). Reference and attitude in infant pointing. *Journal of Child Language*, 34(1), 1-20.
- Liszkowski, U., Carpenter, M., & Tomasello, M. (2007b). Pointing out new news, old news, and absent referents at 12 months of age. *Developmental Science*, 10(2), F1-F7.
- Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, 108(3), 732-739.
- Locke, J.L. (1983). *Phonological Acquisition and Change*, Academic Press, New York.
- Loeb, D.F. & Allen, G.D. (1993). Preschoolers' imitation of intonation contours. *Journal of Speech and Hearing Research* 36, 4–13.
- 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.
- MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning & Verbal Behavior*, 23(2), 127-150.
- Maratsos, M. P. & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In Keith E. Nelson (ed.), *Children's language*, vol. 2, New York: Gardner Press, 127–214.
- Marcus, G. F., Vijayan, S., Bandi Rao, S. & Vishton, P.M. (1999). Rule learning by seven-month-old-infants. *Science* 283, 77-80.

- Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins and domain specificity. In M. R. Gunnar & M. P. Maratsos (Eds.), *Modularity and constraints in language and cognition*, Hillsdale, NJ: Erlbaum, 59-101.
- Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, 20, 121–157.
- Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, 47, 241–275.
- Matthews, D., Theakston, A., Lieven, E. & Tomasello M. (2009). Pronoun co-referencing errors: challenges for generativist and usage-based accounts. *Cognitive Linguistics*, 2, 599-626.
- Mattys, S. L. & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78. 91–121.
- Mehler, J., Bertoncini, J. & Barrière, M. (1978). Infant recognition of mother's voice, *Perception* 7, 491–497.
- Meltzoff, A. N., & Brooks, R. (2007). Eyes wide shut: The importance of eyes in infant gaze following and understanding other minds. In: R. Flom, K. Lee, & D. Muir (Eds.), *Gaze following: Its development and significance*. Mahwah, NJ: Erlbaum, 217-241.
- Merriman, W. E., & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, 54(3–4) (Serial No. 220) 1–129.
- Miller, W. & Ervin, S. (1964). The development of grammar in child language. In U. Bellugi & R. Brown (eds), *The acquisition of language*. Monogr. Soc. Res. Ch. Devel. 29.
- Mintz, T. H., Newport, E. L. & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science* 26. 393–424.
- 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. (2007). How 14- and 18-month-olds know what others have experienced. *Developmental Psychology*. 43(2), 309-317.

- Moll, H., Carpenter, M., & Tomasello, M. (2007). Fourteen-month-old infants know what others experience only in joint engagement with them. *Developmental Science*, 10(6), 826-835.
- Moore, C., & D'Entremont, B. (2001). Developmental changes in pointing as a function of attentional focus. *Journal of Cognition & Development*, 2(2), 109-129.
- Morgan, J. L. (1986). *From simple input to complex grammar*. Cambridge, MA: MIT Press.
- Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, Vol. 111, pp. 855-69.
- Müller, A., Höhle, B., Schmitz, M., & Weissenborn, J. (2009). Information structural constraints on children's early language production: The acquisition of the focus particle auch ('also') in German-learning 12- to 36-month-olds. *First Language*, 29(4), 373-399.
- Nespor, M., Guasti, M. T., & Christophe, A. (1996). Selecting word order: the Rhythmic Activation Principle. In U. Kleinhenz (Ed.), *Interfaces in Phonology* (pp. 1-26). Berlin: Akademie Verlag.
- Newman (1946). On the stress system of English, *Word* 2, 171-187.
- Ohala, J. J. (1980) The Acoustic Origin of the Smile. *Journal of the Acoustical Society of America* 68, 33.
- Ohala, J. J. (1983). CrossLanguage Use of Pitch: An Ethological View. *Phonetica* 40, 1-18.
- O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, 67, 659-677.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, 308(5719), 255-258.
- Öhman, S. (1967). Word and sentence intonation: a quantitative model. *Speech Transmission Laboratory Quarterly Progress and and Status Report* 2-3:20-54.
- Papousek, M., Papousek, H., Haekel, M. (1987). Didactic adjustments in fathers' and mothers' speech to their 3-month-old infants. *Journal of Psycholinguistic Research*, 16, 491-516.

- Pelzer, L. & Höhle, B. (2006). Processing of morphological markers as a cue to syntactic phrases by 10-month-olds German-learning infants. In Adriana Belletti, Elisa Bennati, Cristiano Chesi, Elisa DiDomenico & Ida Ferrari. (eds.), *Language acquisition and development: Proceedings of GALA 2005*, Cambridge: Cambridge Scholars Press, 411–422.
- Perner, J. & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, 308(5719), 214-216.
- 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*. Cambridge: MIT Press. 271-311.
- Pine, J. M., & Lieven, E. (1997). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics*, 18(2), 123-138.
- Pinker, St. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pinker, St. (1987). The Bootstrapping Problem in Language Acquisition. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum, 399-441.
- Pinker, St. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.
- Plutchik, R. (2001). The nature of emotions. *American Scientist* (89). 344-350
- deRuiter, L. (2010). *Studies on intonation and information structure in child and adult german*. PhD thesis, Max Planck Institute for Psycholinguistics, Nijmegen.
- Rooth, M. (1992). A Theory of Focus Interpretation. *Natural Language Semantics* 1, 75-116.
- Saffran, J. R., Aslin R. & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science* 274. 1926–1928.
- Salomo, D., Lieven, E., & Tomasello, M. (2010). Young children's sensitivity to new and given information when answering predicate-focus questions. *Applied Psycholinguistics*, 31(1), 101-115.
- Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter and Tomasello. *Child Development*, 1, 94-104.

- Saylor, M. M., Sabbagh, M. A., & Baldwin, D. A. (2002). Children use whole-part juxtaposition as a pragmatic cue to word meaning. *Developmental Psychology*, 38(6), 993-1003.
- Saylor, M. M., Baldwin, D. A., & Sabbagh, M. A. (2004). Converging on word meaning. In: D. G. Hall & S. R. Waxman (Eds.). *Weaving a lexicon*. Cambridge, MA: MIT Press, 509-531.
- Schafer, A.J., Speer, S.R., Warren, P. & White, D. (2000) Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research* 29, 169-182.
- Scherer, K. R. (2003). Vocal communication of emotion, *Speech and Communication*, 40(1-2), 227–256.
- Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*, Cambridge, Eng.: Cambridge University Press.
- Selkirk, E. (1984). *Phonology and Syntax. The Relation between Sound and Structure*. Cambridge, MA: MIT Press.
- Shwe, H. I. & Markman, E. M. (1997). Young children's appreciation of the mental impact of their communicative signals. *Developmental Psychology* 33, 630-636.
- Silverman, K. (1987). *The Structure and Processing of Fundamental Frequency Contours*. PhD thesis, University of Cambridge.
- Skinner, B. F. (1957). *Verbal behavior*. New York, NY: Appleton-Century-Crofts.
- Sluifster, A. M. (1995). *Phonetic correlates of stress and accent*. Dissertation, University of Leiden.
- Snedeker, J. & Yuan, S. (2008) Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language* 58, 574-608.
- 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), 576-87.
- Speer, S. R., Warren, P. & Schafer, A. J. (2003). Intonation and sentence processing. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona 2003*. Rundle Mall: Causal Productions, 95-105.
- Sperber, D. & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.) Oxford: Blackwell.

- Stoll, S. (1998). The Role of Aktionsart in the Acquisition of Russian Aspect. *First Language*, 18, 351-378.
- Szagan, G, Stumper, B. & Schramm, S.A. (2009). *Fragebogen zur frühkindlichen Sprachentwicklung (FRAKIS) und FRAKIS-K (Kurzform)*. Frankfurt: Pearson Assessment.
- Taylor, J. R. (2002). *Cognitive Grammar*. Oxford, Oxford University Press.
- Taylor, P.A. (1992). *A phonetic model of English intonation*. PhD Thesis, University of Edinburgh (1992) (published by Indiana University Linguistics Club).
- Thorsen, N. (1979a). Lexical stress, emphasis for contrast and sentence intonation in Advanced Standard Copenhagen Danish, ARIPUC 13, 59-85.
- Tomasello, M. (1992). *First Verbs: A Case Study of Early Grammatical Development*. Cambridge University Press.
- Tomasello, M. (1995a). Joint attention as social cognition. In: C. Moore and P.J.Dunham (eds.). *Joint attention: Its origins and role in development*. Hillsdale, NJ: Erlbaum.
- Tomasello, M. (1998a). *The new psychology of language*, vol. 1: *Cognitive and functional approaches to language structure*. Mahwah, NJ: Erlbaum
- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74(3), 209-253.
- Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. In M. Bowerman & S. Levinson (Eds.), *Language Acquisition and Conceptual Development*. Cambridge University Press.
- Tomasello, M. (2003). *Constructing a Language. A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2008). *Origins of Human Communication*. MIT Press.
- Tomasello, M., & Barton, M. E. (1994). Learning words in nonostensive contexts. *Developmental Psychology*, 30(5), 639-650.
- Tomasello, M. & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development*, 10, 201-224.
- Tomasello, M. & Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what's new for other persons. *Developmental Psychology*, 39, 906-912.

- Turk, A. E., & White, L. (1999). Structural influences on accentual lengthening. *Journal of Phonetics* 27, 171–206.
- 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.
- Vihman, M. & Croft, W. (2007). Phonological development: Toward a 'radical' templatic phonology. *Linguistics* 45, 683-725.
- Vogel, I. & Raimy, E. (2002). The acquisition of compound vs. phrasal stress: the role of prosodic constituents. *Journal of Child Language* 29, 225-250.
- Warren, P., Schafer, A.J., Speer, S.R., & White, S.D. (2000). Prosodic resolution of prepositional phrase ambiguity and unambiguous situations. *UCLA Working Papers in Phonetics*, 99: 5-33.
- Weber, A., Grice, M. & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: a study of anticipatory eye movements. *Cognition* 99(2), B63-B72.
- Wells, B., Peppe, S. & Goulandris, N. (2004). Intonation development from five to thirteen, *Journal of Child Language* 31 (2004), 749–778.
- Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology*, 43, 230-246.
- Werker, J. F., Pegg, J. E., & McLeod, P. J. (1994). A cross-language investigation of infant preference for infant-directed communication. *Infant Behavior and Development*, 17, 323–333.
- Wiemann, L. A. (1976): Stress patterns in early child language. *Journal of Child Language* 3, 283-286.
- Woodward, A. L. & Markman, E. M. (1998). Early word learning. In: W. Damon, D. Kuhn & R. Siegler, (Eds.) *Handbook of child psychology, Volume 2: Cognition, perception and language*. New York: John Wiley and Sons, 371-420.
- Xu, Y. & Xu, C. X. (2005) Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33, 159–197.

12. Appendix

Appendix A: Test sentences 'Resolving syntactic ambiguities' (Chapter 7.1.)

Case Marking / Contrastive Intonation condition

Den Papagei wieft der Löwe.

The (acc-masc) parrot is weefing the (nom-masc) lion.

Den Tiger tammt der Frosch.

The (acc-masc) tiger is tamming the (nom-masc) frog.

Den Pinguin bafft der Fisch..

The (acc-masc) penguin is baffing the (nom-masc) fish..

Den Hahn mommelt der Eisbär.

The (acc-masc) cock is mommeling the (nom-masc) ice bear.

Case Marking / Neutral Intonation condition

Den Hund wieft der Elefant.

The (acc-masc) dog is weefing the (nom-masc) elephant.

Den Bär tammt der Affe.

The (acc-masc) bear is tamming the (nom-masc) ape.

Den Gorilla bafft der Hase.

The (acc-masc) gorilla is baffing the (nom-masc) rabbit.

Den Igel mommelt der Hirsch.

The (acc-masc) hedgehog is mommeling the (nom-masc) deer.

No Case Marking / Contrastive Intonation condition

Die Kuh wieft die Maus.

The (ambiguous-fem) cow is weefing the (ambiguous-fem) mouse.

Die Ziege mommelt die Spinne.

The (ambiguous-fem) goat is mommeling the (ambiguous-fem) spider.

Das Zebra tammt das Eichhörnchen.

The (ambiguous-neuter) zebra is tamming the (ambiguous-neuter) squirrel.

Das Krokodil bafft das Huhn.

The (ambiguous-neuter) crocodile is baffing the (ambiguous-neuter) chicken..

No Case Marking / Neutral Intonation condition

Die Katze bafft die Gans.

The (ambiguous-fem) cat is baffing the (ambiguous-fem) goose.

Die Schlange tammt die Giraffe.

The (ambiguous-fem) snake is tamming the (ambiguous-fem) giraffe.

Das Schwein wieft das Pferd.

The (ambiguous-neuter) pig is weefing the (ambiguous-neuter) horse.

Das Schaf mommelt das Erdmännchen.

The (ambiguous -neuter) sheep is mommeling the (ambiguous -neuter) meerkat.

Appendix B: Test sentences 'The role of context & intonation in resolving syntactic ambiguities' (Chapter 7.2.)

Case Marking / Contrastive Intonation condition

P1: Ich glaube, der Löwe wieft den Frosch!
I think, the lion (nom-masc) is weefing the frog (acc-masc)!

P2: Nicht den Löwen wieft der Frosch, sondern den Papagei wieft der Löwe.
It's not the lion (acc-masc) that's weefing the frog, it's the parrot (acc-masc) that's weefing the lion.

P1: Ich glaube, der Frosch tammt den Pinguin!
I think, the frog (nom-masc)is weefing the penguin (acc-masc)!

P2: Nicht den Pinguin wieft der Frosch, sondern den Tiger tammt der Frosch.
It's not the penguin (acc-masc) that's weefing the frog (nom-masc), it's the (acc-masc) tiger that's tamming the (nom-masc) frog.

P1: Ich glaube, der Fisch bafft den Tiger!
I think, the fish (nom-masc) is baffing the tiger (acc-masc)!

P2: Nicht den Tiger bafft der Fisch, sondern den Pinguin bafft der Fisch.
It's not the tiger (acc-masc) that's baffing the fish (nom-masc), it's the (acc-masc) penguin that's baffing the (nom-masc) fish.

P1: Ich glaube, der Eisbär mommelt den Esel!
I think, the ice bear (nom-masc)is mommeling the donkey (acc-masc)!

P2: Nicht den Esel mommelt der Eisbär, sondern den Hahn mommelt der Eisbär.
It's not the donkey (acc-masc) that's mommeling the ice bear (nom-masc), it's the (acc-masc) cock that's mommeling the (nom-masc) ice bear.

Case Marking / Neutral Intonation condition

P1: Ich glaube, der Elefant wieft den Papagei!

I think, the elephant (nom-masc) is weefing the parrot (acc-masc)!

P2: Nicht den Papagei wieft der Elefant, sondern den Hund wieft der Elefant.

It's not the parrot (acc-masc) that's weefing the elephant (nom-masc), it's the (acc-masc) dog that's weefing the (nom-masc) elephant.

P1: Ich glaube, der Affe tammt den Hahn!

I think, the ape (nom-masc) is tammng the cock (acc-masc)!

P2: Nicht den Hahn tammt der Affe, sondern den Bär tammt der Affe.

It's not the cock (acc-masc) is tammng the ape (nom-masc), it's the (acc-masc) bear that's tammng the (nom-masc) ape.

P1: Ich glaube, der Hase bafft den Koala!

I think, the rabbit (nom-masc) is baffng the koala (acc-masc)!

P2: Nicht den Koala bafft der Hase, sondern den Gorilla bafft der Hase.

It's not the koala (acc-masc) that's baffng the rabbit (nom-masc), it's the (acc-masc) gorilla that's baffng the (nom-masc) rabbit.

P1: Ich glaube, der Hirsch mommelt den Adler!

I think, the deer (nom-masc) is mommelng the eagle (acc-masc)!

P2: Nicht den Adler mommelt der Hirsch, sondern den Igel mommelt der Hirsch.

It's not the eagle (acc-masc) that's mommelng the deer (nom-masc), it's the (acc-masc) hedgehog that's mommelng the (nom-masc) deer. 42

No Case Marking / Contrastive Intonation condition

P1: Ich glaube, die Maus wieft die Spinne!

I think, the mouse (ambiguous-fem) is weefing the spider (ambiguous-fem)!

P2: Nicht die Spinne wieft die Maus, sondern die Kuh wieft die Maus.

It's not the spider (ambiguous-fem) that's weefing the mouse (ambiguous-fem), it's the (ambiguous-fem) cow that's weefing the (ambiguous-fem) mouse

P1: Ich glaube, das Eichhörnchen tammt das Schwein!

I think, the squirrel (ambiguous-neuter) is tamming the pig (ambiguous-neuter)!

P2: Nicht das Schwein tammt das Eichhörnchen, sondern das Zebra tammt das Eichhörnchen.

It's not the pig (ambiguous-neuter) that's tamming the squirrel (ambiguous-neuter), it's the (ambiguous-neuter) zebra that's tamming the (ambiguous-neuter) squirrel.

P1: Ich glaube, das Huhn bafft das Erdmännchen!

I think, the chicken (ambiguous-neuter) is baffing the meerkat (ambiguous-neuter)!

P2: Nicht das Erdmännchen bafft das Huhn, sondern das Krokodil bafft das Huhn.

It's not the meerkat (ambiguous-neuter) that's baffing the chicken (ambiguous-neuter), it's the (ambiguous-neuter) crocodile that's baffing the (ambiguous-neuter) chicken.

P1: Ich glaube, die Spinne mommelt die Schlange!

I think, the spider (ambiguous-fem) is mommeling the snake (ambiguous-fem)!

P2: Nicht die Schlange mommelt die Spinne, sondern die Ziege mommelt die Spinne.

It's not the snake (ambiguous-fem) that's mommeling the spider (ambiguous-fem), it's the (ambiguous-fem) goat that's mommeling the (ambiguous-fem) spider.

No Case Marking / Neutral Intonation condition

P1: Ich glaube, das Pferd wieft das Krokodil!

I think, the horse (ambiguous-neuter) is weefing the crocodile (ambiguous-neuter)!

P2: Nicht das Krokodil wieft das Pferd, sondern das Schwein wieft das Pferd.

It's not the crocodile (ambiguous-neuter) that's weefing the horse (ambiguous-neuter), it's the (ambiguous-neuter) pig that's weefing the (ambiguous-neuter) horse.

P1: Ich glaube, die Giraffe tammt die Ziege!

I think, the giraffe (ambiguous-fem) is tamming the goat (ambiguous-fem)!

P2: Nicht die Ziege tammt die Giraffe, sondern die Schlange tammt die Giraffe.

It's not the goat (ambiguous-fem) that's tamming the giraffe (ambiguous-fem), it's the (ambiguous-fem) snake that's tamming the (ambiguous-fem) giraffe.

P1: Ich glaube, die Gans bafft die Giraffe!

I think, the goose (ambiguous-fem) is baffing the giraffe (ambiguous-fem)!

P2: Nicht die Giraffe bafft die Gans, sondern die Katze bafft die Gans.

It's not the giraffe (ambiguous-fem) that's baffing the goose (ambiguous-fem), it's the (ambiguous-fem) cat that's baffing the (ambiguous-fem) goose

P1: Ich glaube, das Erdmännchen mommelt das Huhn!

I think, the meercat (ambiguous-fem) is mommeling the chicken (ambiguous-fem)!

P2: Nicht das Huhn mommelt das Erdmännchen, sondern das Schaf mommelt das Erdmännchen.

It's not the chicken (ambiguous-fem) that's mommeling the meercat (ambiguous-fem), it's the (ambiguous -neuter) sheep that's mommeling the (ambiguous -neuter) meercat.

Appendix C: Picture books: 'Young children's intonational marking of new and given referents' (Chapter 8) & 'The role of the input for children's intonational development' (Chapter 9)

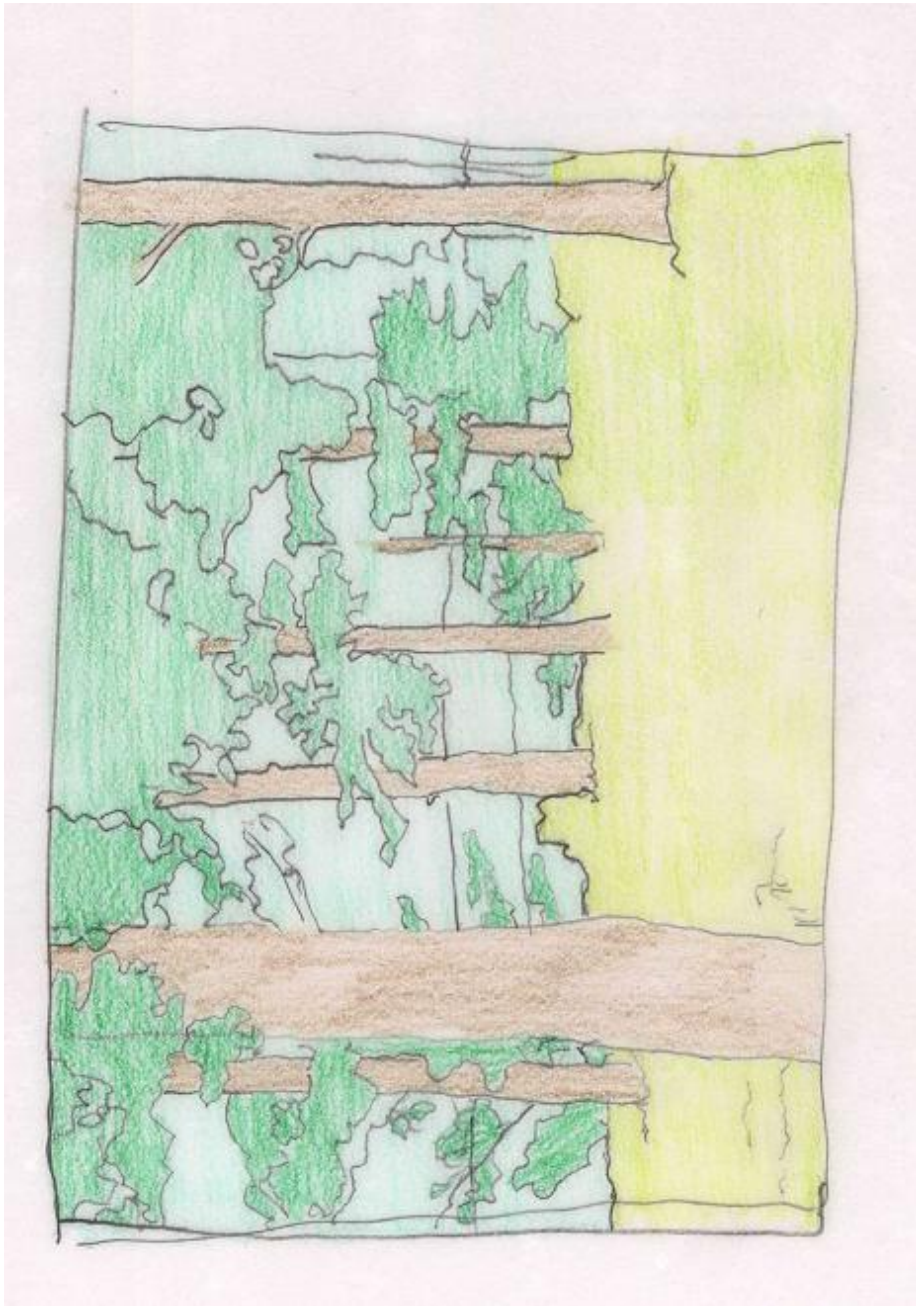


Figure A: Example of the first picture of the picture-books. The picture was intended to introduce the topic (e.g. a forest).



Figure B: Example of the second picture of the picture-books. Picture 2 was intended to introduce the target referent (e.g. a hedgehog)



Figure C: Example of the third picture of the picture-books. The picture was intended to introduce a distractor referent (e.g. a deer). In order to keep the target referent active, the target referent was visible in the background of the picture.



Figure D: Example of the fourth picture of the picture-books. The picture shows the distractor referent acting on the target referent in a causative way (e.g. the deer is washing the hedgehog). The picture attempted to elicit a transitive SVO sentence, in which the target referent was mentioned as the patient.



Figure E: Example of the fifth picture of the picture-books. The picture shows the distractor referent acting on the target referent in a causative way (e.g. the deer is combing the hedgehog). The picture attempted to elicit a transitive SVO sentence, in which the target referent was mentioned as the patient.



Figure F: Example of the sixth picture of the picture books. The picture shows how the target referent left the scene. The picture attempted to elicit a contrastive utterance (as response or protest to the experimenters wrong naming of the target referent).

Appendix D: Examples of utterances ‘Young children’s intonational marking of new and given referents’ (Chapter 8) & ‘The role of the input for children's intonational development’ (Chapter 9)

	new	given	contrastive
2;6	die biene ist gekommen "the bee is come" <i>The bee has come</i>	der grashüpfer tut die biene in arm nehmen "the grasshopper does the bee in arm take" <i>The gr. hugs the bee</i>	ne, das war die biene "no, that was the bee" <i>no, that was the bee</i>
3;0	'ne eule sitzt da drin "an owl sits there in" <i>an owl is sitting in there</i>	der guckt nach die eule "he looks to the owl" <i>He is looking after the owl"</i>	nein, eine eule, kein fisch "no, an owl, no fish" <i>No, an owl, not a fish</i>
adult	jetzt kommt eine möwe auf diesen strand zugeflogen "now comes a seagull to this beach flown" <i>now a seagull is flying towards this beach</i>	jetzt ist die möwe scheinbar gelandet "now is the seagull apparently landed" <i>now the seagull apparently landed</i>	ne, sie streut vor die möwe "no, she strews in front of the seagull" <i>no, she strews in front of the seagull</i>
CDS	und da kommt ein kleiner igel "and there comes a small hedgehog" <i>and there comes a small hedgehog"</i>	der hirsch der kämmt die borsten vom igel "the deer he combs the bristles of the hedgehog" <i>the deer he combs the bristles of the hedgehog</i>	

Figure G: The diagram shows examples of the utterances that participants from each groups (2;6 years, 3;0 years, adults and CDS) gave in each of the three conditions (‘new’, ‘given’ and ‘contrastive’). The original utterance is printed in bold, the loose translation in inverted commas. Finally, a grammatical translation is shown in italics.

